

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOTEHNOLOOGIA ÕPPETOOL

**Sekkumist võimaldavad geenivariandid TÜ Eesti Geenivaramu valimis –
haigusseoselisuse tuvastamine ja võimalik kasutamine geneetilises
tagasisides**

Magistritöö (30 EAP)

Bioloogia

Karmen Vaiküll

Juhendajad PhD Tiit Nikopensius
MD, PhD Neeme Tõnisson

TARTU 2018

INFOLEHT

Sekkumist võimaldavad geenivariandid TÜ Eesti Geenivaramu valimis – haigusseosisuse tuvastamine ja võimalik kasutamine geneetilises tagasisides

Täisgenoomi ja eksoomi sekveneerimine võimaldab tuvastada uuritavatel kõrge haigusriskiga geenivariante, mille kohta tagasisidet andes oleks võimalik haigusest tulenevaid komplikatsioone ära hoida. Käesoleva magistritöö eesmärgiks oli otsida Tartu Ülikooli Eesti Geenivaramu geenidonoritel võimalikke sekkumistõhusaid geenivariante, et hinnata käsitletud haiguste esinemissagedust Eestis ning võimaldada tulevikus haigusriskide osas ka tagasisidet. Töös hinnati variante geenides, mis on välja toodud konsensusnimekirjades või seoses lüsosomaalsete ladestushaigustega. Leiti 125 haigusseoselist geenivarianti 363 isikul, kellest 120-l võib olla kõrge haigusrisk ning kes võiksid tulevikus potentsiaalselt tagasisidet saada.

Märksõnad: sekkumistõhusad geenivariandid, geneetiline tagasiside

CERCS: B110 bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Actionable gene variants in Estonian Genome Center sample – identification of disease-associated variants and their possible use in genetic feedback

Whole-genome and exome sequencing allows identification of high risk genetic variants and by giving genetic feedback, it is possible to prevent the complications of the disease. The aim of this study was to search for actionable gene variants in Estonian Genome Center gene donors, assess the incidence of addressed diseases in Estonia and to provide the return of genomic results to gene donors. Variants evaluated in this study were in genes named in actionable gene lists or related to lysosomal storage disorders. 125 disease related gene variants were discovered in 363 individuals, of whom 120 could potentially have a high disease risk and should receive feedback in the future.

Keywords: actionable gene variants, genetic feedback

CERCS: B110 bioinformatics, medical informatics, biomathematics, biometrics

SISUKORD

KASUTATUD LÜHENDID	5
SISSEJUHATUS	6
1. KIRJANDUSE ÜLEVAADE.....	7
1.1. Geneetilised uuringud teaduses	7
1.2. Geneetilise tagasiside andmine.....	7
1.3. Sekkumist võimaldavad geneetilised haigused, geenide konsensusnimekirjad .	9
1.4. Lüsosomaalsed ladestushaigused	13
1.5. Haigusseoselisuse hindamine	15
1.5.1. Kliinilised andmebaasid	16
1.5.1.1. HGMD	16
1.5.1.2. ClinVar	17
1.5.1.3. InterVar.....	17
1.5.1.4. Teised andmebaasid.....	18
2. EKSPERIMENTAALOSA	19
2.1. Töö eesmärgid	19
2.2. Materjal ja metoodika.....	19
2.2.1. Valim	19
2.2.2. Sekveneerimisandmed	20
2.2.3. Variantide prioritseerimine.....	20
2.2.4. Alleelisageduse referentsandmebaasid	20
2.2.5. <i>In silico</i> skoorid	21
2.2.6. Kasutatud andmebaasid	22
2.2.7. Arvutused	23
2.2.8. Lüsosomaalsete ladestushaigustega seotud geenid	23
2.3. Tulemused ja arutelu	24
2.3.1. ACMG/Geisinger nimekirjaga seotud tulemused.....	24
2.3.2. Lüsosomaalsete ladestushaigustega seotud tulemused.....	33
KOKKUVÕTE	42
SUMMARY	43
TÄNUAVALDUSED.....	44

KASUTATUD KIRJANDUSE LOETELU	45
KASUTATUD VEEBIAADRESSID	51
LISA1	52
LISA 2	55
Lihtlitsents	57

KASUTATUD LÜHENDID

ACMG	Ameerika Meditsiinigeneetika ja -genoomika Kolleegium, <i>American College of Medical Genetics and Genomics</i>
GHS	Geisingeri Tervishoiu Süsteem, <i>Geisiger Health System</i>
NGS	järgmise põlvkonna sekveneerimine, <i>next generation sequencing</i>
LSD	lüsomaalsed ladestushaigused, <i>lysosomal storage diseases</i>
ERT	ensüümasendusravi, <i>enzyme replacement therapy</i>
MPS	mukopolüsahharidoos
WGS	täisgenoomi sekveneerimine, <i>whole-genome sequencing</i>
WES	täiseksoomi sekveneerimine, <i>whole-exome sequencing</i>
AMP	Molekulaarpatoloogia Assotsiatsioon, <i>Association for Molecular Pathology</i>
VEP	<i>Variant Effect Predictor</i>
HGMD	<i>Human Gene Mutation Database</i>
HGVS	<i>Human Genome Variation Society</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
LOVD	<i>Leiden Open Variation Database</i>
TÜ EGV	Tartu Ülikooli Eesti Geenivaramu
gnomAD	<i>The Genome Aggregation Database</i>
1000G	<i>1000 Genomes Project</i>
ESP	<i>Exome Sequencing Project</i>
ESE	eksonis paiknev splaissingu võimendaja, <i>exonic splicing enhancer</i>
ESS	eksonis paiknev splaissingu vaigistaja, <i>exonic splicing silencer</i>

SISSEJUHATUS

Kaasaja meditsiinilises geneetikas on järjest laienemas eksoomi ja genoomi sekveneerimine ning seetõttu on võimalik tuvastada uuritavatel indiviididel rohkem geneetilisi variante, millel võib olla kliiniline tähtsus. Sellest tulenevalt on hakatud tähelepanu pöörama geneetilise tagasiside andmisele, et võimaldada inimestele õigeaegset diagnoosi ja ravi.

Selle tarbeks on loodud nimekirju sekkumistõhusatest geenidest, milles olevate haigusseoseliste mutatsioonide tuvastamise korral tuleks tagasisidet anda. Sekkumist võimaldavateks haigusteks on näiteks erinevad pärilikud vähi sündroomid ja südamehaigused. Kiirelt arenevate ravimeetodite kättesaadavuse tõttu on muutunud potentsiaalselt sekkumistõhusaks ka lüsosomaalsed ladestushaigused, mida põhjustavad mutatsioonid lüsosomaalseid ensüüme kodeerivates geenides.

Tagasiside oluliseks küsimuseks on geenivariantide haigusseoselisuse hindamine, milleks on loodud erinevaid kliinilisi andmebaase ja analüütilisi algoritme, mis aitavad kliinilist olulisust tõlgendada.

Seni on avaldatud mitmeid teadusuuringuid, kus on küsitletud, kas inimesed soovivad geneetiliste leidude kohta tagasisidet. Mitmetes riikides ja keskustes on nüüdseks käivitatud genoomipõhise tagasiside projekte ning inimestel on võimalik geneetilise nõustamise vormis tagasisidet saada.

Käesoleva magistriöö eesmärgiks on ACMG (*American College of Medical Genetics and Genomics*) (Kalia jt., 2016) ning Geisinger-76 (Dewey jt., 2016) nimekirjade alusel otsida Tartu Ülikooli Eesti Geenivaramu geenidonoritel võimalikke sekkumistõhusaid geenivariante, hinnata vastavate haiguste eeldatavat esinemissagedust Eestis ning võimaldada tulevikus inimestele kõrgete haigusriskide osas ka tagasisidet.

Töö teiseks eesmärgiks on hinnata lüsosomaalsete ladestushaiguste eeldatavat sagedust Eestis ning otsida potentsiaalseid kõrge haigusriskiga isikuid, kellel võib esineda haiguse hilise avaldumisega vorme.

1. KIRJANDUSE ÜLEVAADE

1.1. Geneetilised uuringud teaduses

Kuigi viimastel aastatel on leitud suure hulga erinevate genoomi piirkondade seos haigusriskidega, on enamuse geneetiliste variantide seosed haigustega veel tundmatud ning uuringud jätkuvad (Durbin jt., 2010). Üks paremaid viise selliseid uuringuid läbi viia on populatsioonipõhised biopangad, mis võimaldavad suurte andmehulkade kättesaadavust. Populatsioonipõhised biopangad sisaldavad endas tuhandete inimeste geeniproove ning lisaks genealoogilisi, meditsiinilisi ja elustiiliga seotud andmeid, mida korrapäraselt uuendatakse (Parodi, 2015). Sidudes meditsiinilisi ja elustiili andmeid geneetikaga, on võimalik tuvastada haigustega seotud variante, samuti saab populatsioonipõhise biopanga andmestiku põhjal hinnata alleelide sagedust populatsioonis (Swede jt., 2007). Haigusi, mida põhjustab selgelt üks geenimutatsioon, esineb harva ning nende puhul on geneetiliste ja keskkonnategurite komponenti lihtsam eristada kui tavaliste komplekshaiguste puhul nagu diabeet, Alzheimeri tõbi, astma, skisofreenia ja pahaloomulised kasvaja (Swede jt., 2007). Monogeensete ja komplekshaiguste geneetiliste põhjuste uurimist võimaldavad järgmise põlvkonna sekveneerimise (NGS, *next generation sequencing*) tehnoloogiad, mille abil on võimalik järjestada kogu genoom või eksoom (Kulski, 2016). NGS on ulatuslikum kui suunatud resekveneerimine ja geenikiibid (Sanderson jt., 2017) ning võimaldab avastada mutatsioone, millel võib olla kliinilisi seosed (Blackburn jt., 2015).

Populatsioonipõhiste biopankade üheks eesmärgiks on panna alus uute biomeditsiiniliste teadmiste loomisele ning välja töötada viise kuidas teadmisi praktikas kasutada (Murtagh jt., 2011). Geneetilistest uuringutest tulenevad andmed muudavad õigeaegse diagnoosimise lihtsamaks ning võimaldavad määrata sobiva ravi. Geenianalüüsi uurimine on üks olulisemaid uusi arengusuundi meditsiini valdkonnas.

Tartu Ülikooli Eesti Geenivaramu on Euroopas üks edukamaid populatsioonipõhiseid biopanku, mis sisaldab hetkel rohkem kui 52 000 geenidoonori andmeid ning võimaldab inimgeeni-uuringute seadusest lähtuvalt rakendada geenianalüüsi rahvatervise parandamiseks.

1.2. Geneetilise tagasiside andmine

Kuna DNA sekveneerimise tehnoloogiad võimaldavad kliiniliselt oluliste variantide tuvastamist, oleks võimalik sellest inimestele teada andes ja nõustades potentsiaalselt haigust ennetada ja kontrollida. Kokkuvõttes võib see kaasa tuua teatud haiguste varajasema avastamise ning suurendada tervena elatud aastate arvu. Seetõttu on olnud suur huvi haigusseoseliste variantide tulemuste tagastamise suuniste väljatöötamiseks (Johnson ja

Gehlert, 2014). Geneetiliste tulemuste tagasiside on üldiselt kujunenud aga keeruliseks küsimuseks, mille üle arutletakse pidevalt. Sellele vaatamata on mitmel pool kokku lepitud, et geneetilisi tulemusi peaks vastavalt heakskiidetud juhiste tagastama (Johns jt., 2017). Arutelude põhjuseks on see, et näiliselt tervetele indiviididele tulemuste tagastamine kogu genoomi sekveneerimise tulemusena võib küll parandada heaolu läbi suurenenud haigestumiskirguse prognoosimise, ennetamise ja diagnoosimise, kuid tekitab ka mitmeid praktilisi ja eetilisi probleeme (Sanderson jt., 2017).

Üheks aspektiks on see, et teadusuuringud ei paku kliinilist ravi ja teadlaste ning uuringus osalejate suhe ei ole sama mis arsti ja patsiendi suhe (Johnson ja Gehlert, 2014). Samuti võivad osalejad oma tulemusi valesti tõlgendada või teha halvasti informeeritud ravigratsuseid (Pulford jt., 2016). On leitud, et isegi suhteliselt kõrge haridustasemega indiviidide jaoks on raske mõista, mis on DNA, pärilikkus ja penetrantsus (Bijlsma jt., 2018). Ka arstid peaksid olema tulevikus võimelised kasutama geneetilist informatsiooni tavapärase praktikas, kuid muret geneetilise kirjaoskuse ja valmisoleku kohta on tänasel päeval avaldanud ka nemad ise (Johns jt., 2017).

Kogu genoomi sekveneerimise kasulikkuse hindamisel on oluline mõista, kuidas tulemused võivad indiviide psühholoogiliselt mõjutada. Haruldaste haigustega seotud variandid võivad tekitada inimesetele tugevamaid emotsionaalseid reaktsioone ja seeläbi suuremat motivatsiooni järgnevat samm ette võtta (Sanderson jt., 2017). On tõenäoline, et inimesed ei pruugi soovida teada saada kõike, mida on võimalik tervise kohta geneetilisest andmetest välja lugeda. See tähendab, et indiviididel peab olema võimalus otsustada, kas ja millist geneetilist tagasisidet nad soovivad. Informeeritud nõusolekudokument peab sisaldama nii saadava kasu kui ka võimalike riskide kirjeldust (Johnson ja Gehlert, 2014).

On mitmeid argumente geneetiliste uurimistulemuste tagastamise kasuks, sealhulgas osalejate õigus saada enda kohta olulist teavet ning haigusest tulenevate komplikatsioonide vähendamine või ära hoidmine nii uuringus osalejatel kui nende perekondadel (Fossey jt., 2018; Pulford jt., 2016). Teadusuuringud on näidanud, et enamus indiviide on huvitatud oma geneetiliste tulemuste tagasisidest (Pulford jt., 2016; Bijlsma jt., 2018). Ühes uuringus tagastati tulemusi patogeensete (st haigust põhjustavate) leidude kohta seoses vähiseoseliste sündroomidega ning 7 inimest 11-st tagasisidet saanud tegutses sellele vastavalt, näidates sellega tagasiside tähtsust (Johns jt., 2017).

Geneetiliste tulemuste tagastamise miinimumnõuete kohaselt peaksid tulemused olema valideeritud ja kliiniliselt olulised, kus sekkumine võimaldaks mõjutada ravi või indiviidi suunamist (Pulford jt., 2016). Autosoom-retsessiivsete ja X-liiteliste haiguste puhul antakse

kliiniliselt olulistest variantidest teada siis, kui need esinevad homosügootses, liitheterosügootses või hemisügootses olekus (Dewey jt., 2016).

Sellised uuringud, mis hõlmavad geneetiliste uurimistulemuste tagastamist, on suhteliselt uudsed, kuid eeldatavasti muutuvad need tulevikus sagedasemaks (Fossey jt., 2018). Eesti Geenivaramus on alustanud mitmete tagasisideprojektidega ning geenidoonoritel on võimalus soovi korral saada geneetilist nõustamist. Tagasisidet antakse ka näiteks USAs Geisinger MyCode projekti raames (<https://www.geisinger.org/mycode>).

1.3. Sekkumist võimaldavad geneetilised haigused, geenide konsensusnimekirjad

Genoomi ja eksoomi sekveneerimist kasutatakse üha enam haruldaste haiguste iseloomustamisel, personaalse meditsiini rakendamisel, farmakogenoomikas, sünnieelsel skriiningul ning populatsiooni skriinimisel haiguste riski osas. Kõigil neil juhtudel võib leida sekkumist võimaldavaid geenivariante, millel on meditsiiniline olulisus ning mille avastamine võib tuua patsiendile kasu.

Ameerika Meditsiinigeneetika ja -genoomika Kolleegium (ACMG, *American College of Medical Genetics and Genomics*) on avaldanud miinumumnimekirja (Kalia jt., 2016), kuhu kuulub 59 geeni, milles olevate patogeensete või tõenäoliselt patogeensete geenivariantide korral tuleks leiust uuritavale teada anda. Nimekirja kuulus algselt 56 geeni (Green jt., 2013), kuid sealt eemaldati üks ning lisati neli geeni. Nimekiri on mõeldud kasutamiseks eelkõige kliinilises kontekstis, kuid koostajad teadvustavad, et soovitusi võidakse üle võtta ka teadusuuringute raames.

Nimekirja kuuluvad geenid on seotud harvaesinevate haigustega, millel on olemas ravi ja/või ennetusmeetmed ning mille korral haigusseoselised mutatsioonid võivad olla pikka aega asümptomaatilised. Põhiliselt on nimekirja geenid seotud pärilike vähi sündroomide ja südamehaigustega, aga ka näiteks aneurüsme soodustavate sidekoehaigustega ja maliigse hüpertermia eelsoodumusega. Keskendutakse haigustele, mida põhjustavad punktmutatsioonid ning väikesed insertioonid ja deletsioonid (indelid). Tagasisidet soovitatakse anda ainult selliste geenivariantide kohta, mille haigusseoselisusel on piisav toetus kirjanduses ja andmebaasides. ACMG töögrupp soovitab määratletud mutatsioone nimekirja geenides aktiivselt otsida.

Lisaks on loodud Geisingeri Tervishoiusüsteemi (GHS, *Geisiger Health System*) poolt 2016. aastal nn Geisinger-76 nimekiri (GHS76) (Dewey jt., 2016), mis sisaldab ACMG algse 2013. aasta nimekirja 56 geeni ning lisaks 20 geeni, mis on kliinilist sekkumist võimaldavad ja seotud kõrge penetrantsusega monogeensete haigustega, millest enamus on samuti varasemalt ACMG nimekirjas olemas.

GHS76 nimekiri loodi DiscovEHR uuringu raames, mille käigus koostöös Regeneroni Geneetikakeskusega (*Regeneron Genetics Center*) sekveneeriti 50 726 osaleja eksoomid ning seoti andmed elektroonilisest terviseregistrist saadud fenotüübi andmetega. GHS76 nimekirja alusel hinnati uuringus osalejatel potentsiaalselt sekkumist võimaldavate geneetiliste leidude esinemist. GHS76 nimekiri koos ACMG uuendustega on toodud tabelis 1.

Nimekirju sekkumistõhusate geenide kohta on loodud veel, näiteks “Next Medicine” projekti raames loodud 114 geeni (sh 52 geeni ACMG nimekirja hulgast) sisaldav nimekiri, millega on seotud ainult sellised haigused, mis võivad jääda täiskasvanemas diagnoosimata (Dorschner jt., 2013).

Tabel 1. ACMG ja Geisingeri nimekirjade geenid ja vastavad haigused (Kalia jt., 2016; Dewey jt., 2016)

Geen	Haigus	MIM #	Pärandumine	ACMG/GHS76
<i>BRCA1</i>	Pärilik rinna- ja munasarjavähk	604370	AD	ACMG/GHS76
<i>BRCA2</i>	Pärilik rinna- ja munasarjavähk	612555	AD	ACMG/GHS76
<i>TP53</i>	Li-Fraumeni sündroom	151623	AD	ACMG/GHS76
<i>STK11</i>	Peutz-Jeghers'i sündroom	175200	AD	ACMG/GHS76
<i>MLH1</i>	Lynch'i sündroom	120435	AD	ACMG/GHS76
<i>MSH2</i>	Lynch'i sündroom	120435	AD	ACMG/GHS76
<i>MSH6</i>	Lynch'i sündroom	120435	AD	ACMG/GHS76
<i>PMS2</i>	Lynch'i sündroom	120435	AD	ACMG/GHS76
<i>APC</i>	Perekondlik adenomatoosne polüpoos	175100	AD	ACMG/GHS76
<i>MUTYH</i>	MYH-seotud polüpoos; perekondlik adenomatoosne polüpoos tüüp 2, kolorektaalsed hulgiadenoomid; autosoomretsessiivne pilomatrikoomidega kolorektaalne adenomatoosne polüpoos	132600; 608456	AR	ACMG/GHS76
<i>BMPRIA</i>	Juveniilne polüpoosi sündroom	174900	AD	ACMG
<i>SMAD4</i>	Juveniilne polüpoosi sündroom	174900	AD	ACMG
<i>VHL</i>	Von-Hippel Lindau sündroom	193300	AD	ACMG/GHS76
<i>MEN1</i>	Endokriinsete hulgi kasvavate sündroom tüüp 1	131100	AD	ACMG/GHS76
<i>RET</i>	Endokriinsete hulgi kasvavate sündroom tüüp 2; Perekondlik medullaarne kilpnäärme vähk	171400; 162300; 155240	AD	ACMG/GHS76
<i>PTEN</i>	PTEN hamartoomi tuumori sündroom; Bannayan-Riley-Ruvalcaba sündroom	153480	AD	ACMG/GHS76
<i>RB1</i>	Retinoblastoom	180200	AD	ACMG/GHS76
<i>SDHD</i>	Pärilik paraganglioom-feokromotsütoomi sündroom	168000	AD	ACMG/GHS76
<i>SDHAF2</i>	Pärilik paraganglioom-feokromotsütoomi sündroom	601650	AD	ACMG/GHS76
<i>SDHC</i>	Pärilik paraganglioom-feokromotsütoomi sündroom	605373	AD	ACMG/GHS76

<i>SDHB</i>	Pärilik paraganglioom-feokromotsütoomi sündroom	115310	AD	ACMG/GHS76
<i>TSC1</i>	Tuberoosne skleroos	191100	AD	ACMG/GHS76
<i>TSC2</i>	Tuberoosne skleroos	613254	AD	ACMG/GHS76
<i>WT1</i>	WT1-seotud Wilmsi tuumor	194070	AD	ACMG/GHS76
<i>NF2</i>	Neurofibromatoos tüüp 2	101100	AD	ACMG/GHS76
<i>COL3A1</i>	Ehlers-Danlos'i sündroom, vaskulaarne tüüp	130050	AD	ACMG/GHS76
<i>FBN1</i>	Marfan'i sündroom	154700	AD	ACMG/GHS76
<i>TGFBR1</i>	Loeys-Dietz'i sündroom	609192	AD	ACMG/GHS76
<i>TGFBR2</i>	Loeys-Dietz'i sündroom	610168	AD	ACMG/GHS76
<i>TGFB3</i>	Loeys-Dietz'i sündroom	615582	AD	GHS76
<i>SMAD3</i>	Loeys-Dietz'i sündroom	613795	AD	ACMG/GHS76
<i>ACTA2</i>	Perekondlikud rinnaaordi aneurüsmid ja dissekatsioonid	611788	AD	ACMG/GHS76
<i>MYH11</i>	Perekondlikud rinnaaordi aneurüsmid ja dissekatsioonid	132900	AD	ACMG/GHS76
<i>MYBPC3</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	615396; 115197	AD	ACMG/GHS76
<i>MYH7</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	613426; 192600	AD	ACMG/GHS76
<i>TNNT2</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	601494; 115195	AD	ACMG/GHS76
<i>TNNI3</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	613690; 115210; 613286	AD	ACMG/GHS76
<i>TPM1</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	611878; 115196	AD	ACMG/GHS76
<i>ACTC1</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	612098; 613424	AD	ACMG/GHS76
<i>PLN</i>	Hüpertroofiline kardiomiopaatia; Dilatatiivne kardiomiopaatia	613874; 609909	AD	GHS76
<i>MYL3</i>	Hüpertroofiline kardiomiopaatia	608751	AD	ACMG/GHS76
<i>PRKAG2</i>	Hüpertroofiline kardiomiopaatia	600858	AD	ACMG/GHS76
<i>MYL2</i>	Hüpertroofiline kardiomiopaatia	608758	AD	ACMG/GHS76
<i>LMNA</i>	Dilatatiivne kardiomiopaatia	115200	AD	ACMG/GHS76
<i>DES</i>	Dilatatiivne kardiomiopaatia	604765	AD	GHS76
<i>GLA</i>	Fabry tõbi, kardiomiopatiaga	301500	XL	ACMG/GHS76
<i>PKP2</i>	Arütmogeenne parema vatsakese kardiomiopaatia	609040	AD	ACMG/GHS76
<i>DSP</i>	Arütmogeenne parema vatsakese kardiomiopaatia	607450	AD	ACMG/GHS76
<i>DSC2</i>	Arütmogeenne parema vatsakese kardiomiopaatia	610476	AD	ACMG/GHS76

<i>TMEM43</i>	Arütmogeenne parema vatsakese kardiomüopaatia	604400	AD	ACMG/GHS76
<i>DSG2</i>	Arütmogeenne parema vatsakese kardiomüopaatia	610193	AD	ACMG/GHS76
<i>JUP</i>	Arütmogeenne parema vatsakese kardiomüopaatia	611528	AD	GHS76
<i>RYR2</i>	Katehoolaminergiline polümorfne ventrikulaarne tahhükardia	604772	AD	ACMG/GHS76
<i>CAV3</i>	Hüpertroofiline kardiomüopaatia; Pika QT sündroom	192600; 611818	AD	GHS76
<i>KCNQ1</i>	Pika QT sündroom	192500	AD	ACMG/GHS76
<i>KCNH2</i>	Pika QT sündroom	613688	AD	ACMG/GHS76
<i>KCNE1</i>	Pika QT sündroom	613695	AD	GHS76
<i>KCNE2</i>	Pika QT sündroom	613693	AD	GHS76
<i>SNTA1</i>	Pika QT sündroom	612955	AD	GHS76
<i>SCN5A</i>	Pika QT sündroom; Brugada sündroom	603830; 601144	AD	ACMG/GHS76
<i>SCN4B</i>	Pika QT sündroom; Atriaalne fibrillatsioon	611819	AD	GHS76
<i>CACNA1C</i>	Brugada sündroom	611875	AD	GHS76
<i>CACNB2</i>	Brugada sündroom	611876	AD	GHS76
<i>GPD1L</i>	Brugada sündroom	611777	AD	GHS76
<i>HCN4</i>	Brugada sündroom	613123	AD	GHS76
<i>KCNE3</i>	Brugada sündroom	613119	AD	GHS76
<i>SCN1B</i>	Brugada sündroom; Atriaalne fibrillatsioon	612838; 615377	AD	GHS76
<i>SCN3B</i>	Brugada sündroom; Atriaalne fibrillatsioon	613120	AD	GHS76
<i>KCNJ2</i>	Atriaalne fibrillatsioon	613980	AD	GHS76
<i>LDLR</i>	Perekondlik hüperkolesteroleemia	143890	AD	ACMG/GHS76
<i>APOB</i>	Perekondlik hüperkolesteroleemia	144010	AD	ACMG/GHS76
<i>PCSK9</i>	Perekondlik hüperkolesteroleemia	603776	AD	ACMG/GHS76
<i>ATP7B</i>	Wilsoni tõbi	277900	AR	ACMG
<i>RYR1</i>	Maliigse hüpertermia eelsoodumus	145600	AD	ACMG/GHS76
<i>CACNA1S</i>	Maliigse hüpertermia eelsoodumus	601887	AD	ACMG/GHS76
<i>ENG</i>	Pärilik hemorraagiline telangiiektaasia	187300	AD	GHS76
<i>ACVRL1</i>	Pärilik hemorraagiline telangiiektaasia	600376	AD	GHS76
<i>OTC</i>	Ornitiini transkarbamülaasi puudulikkus	311250	XL	ACMG/GHS76

AD – autosoom-dominantne; AR – autosoom-retsessiivne; XL – X-liiteline

MIM # – haiguse identifikaator OMIM (*Online Mendelian Inheritance in Man*) andmebaasis

1.4. Lüsosomaalsed ladestushaigused

Lüsosomaalsed ladestushaigused (LSD, *lysosomal storage diseases*) on pärilikud ainevahetushaigused, mida põhjustavad mutatsioonid lüsosomaalseid valke (sh hüdrolaase ning sünteetil ja transpordil osalevaid valke) kodeerivates geenides (Wang jt., 2011).

Lüsosoom on rakusisene organell, mis vastutab erinevate väliskeskkonnast rakku võetud makromolekulide või raku enda vananenud komponentide lagundamise ning ringlemise võtmise eest. Lüsosoomid sisaldavad üle 60 happelise hüdrolaasi, mis hüdrolyüsivad proteiine, polüsahhariide ja lipiide (Xu ja Ren, 2016). Mutatsioonidest tingitud puudulik lüsosomaalne ensümaatiline aktiivsus ja regulatsioon põhjustavad metaboliseerimata substraadi järk-järgulise akumulierumise lüsosoomi, mis omakorda põhjustab rakkude ja kudede funktsiooni häireid. Näiteks on häiritud raku homöostaas, autofaagia, endo- ja eksotsütoos (Schultz jt., 2011; Rigante jt., 2017). Seega ei põhjusta haigust substraadi lüsosoomi ladestumine iseenesest, vaid sellest tulenevad sekundaarsed efektid, mistõttu on paljudel LSD tüüpidel sarnased kliinilised tunnused, olenemata erinevatest akumulieruvatest makromolekulidest (Rigante jt., 2017).

Lüsosomaalseid ladestushaigusi on ligikaudu 70 ning need moodustavad kollektiivselt 14% kõikidest päritavatest metabolismiga seotud haigustest (Rigante jt., 2017). LSD-sid liigitatakse vastavalt ladestunud substraadi tüübile, nt glükogenoosid (Pompe tõbi), sfingolipidoosid (Fabry, Gaucher', Niemann-Pick'i tõbi), oligosahharidoosid, gangliosidoosid, mukopolüsahharidoosid (MPS) jne (Wang jt., 2011). Kõik LSD-d on harvaesinevad monogeensed haigused. LSD-de kui rühma sagedus on üks juhtum 7000-8000 elussünni kohta (Wang jt., 2011; Rigante jt., 2017). Kõik LSD-d päranduvad autosoom-retsessiivsel viisil, v.a Fabry ja Danoni tõbi ning Hunteri sündroom, mis on X-liitelised.

LSD-d mõjutavad mitmeid organeid (luid, liigeseid, nahka, südant, neere, maksa), põhjustades näiteks organite düsfunktsiooni või suurenemist. Mitmed LSD-d mõjutavad ka kesknärvisüsteemi ja tekitavad neuroloogilisi häireid (Rigante jt., 2017). Haiguste algus, sümptomite raskusaste ja organsüsteemide mõjutatus võivad suuresti erineda, mõnikord isegi peresiseselt (Wang jt., 2011). Kuigi kindlad mutatsioonid võivad olla seotud teatud tagajärgedega, ei ole genotüübi-fenotüübi korrelatsioon alati tugev, nt Gaucher' tõve puhul võib sama mutatsiooni kandlus ühel avalduda juba lapseas, teisel aga ei avaldu terve täiskasvanuea jooksul. X-seoseliste lüsosomaalsete ladestushaiguste puhul määrab naistel haiguse raskuse ja ulatuse peamiselt X-kromosoomi inaktivatsiooni aste (Wang jt., 2011). Kuna LSD-d on haruldased haigused, ei tunne paljud arstid nende sümptomeid ära ning

selleks ajaks kui patsient saab diagnoosi, võivad olla tekkinud pöördumatud kahjustused, mis vähendavad ravi efektiivsust. Paljud patsiendid jäävad aga diagnoosimata (Wang jt., 2011). Patogeneetiline ravi on hetkel olemas vaid loetud lüsosomaalsetele ladestushaigustele (Tabel 2), kuid viimaste aastate jooksul on LSD patsientidele kättesaadavate ravimite arv kiiresti kasvanud ning ravimeetodid arenevad kiiresti. Praegustest meetoditest on saadaval ensüümasendusravi, hematopoeetiliste tüvirakkude siirdamine, substraadi vähendamine ja *chaperon*-teraapiad, kuid arendatakse ka geeniteraapiat ja teisi ravimeetodeid (Beck, 2018). Enimkasutatav on ensüümasendusravi (ERT, *enzyme replacement therapy*) (Wang jt., 2011), mille eesmärgiks on kompenseerida metaboolseid defekte rekombinantsete ensüümide manustamisega. Kasutades spetsiifilist retseptorit, enamasti mannoos-6-fosfaadi retseptorit, on võimalik rakul intravenoosselt manustatavaid ensüüme kätte saada ning transportida neid lüsoosoomidesse, kus ensüümid täidavad oma katalüütilist funktsiooni. LSD ravi aitab stabiliseerida elundite funktsioone või aeglustada haiguse arengut (Beck, 2018).

Tabel 2. Lüsoomaalsed ladestushaigused, millel on ravi olemas või arendamisel

Haigus	Ravi
Gaucher' tõbi	Heaks kiidetud
Fabry tõbi	
MPS I (Hurler, Hurler–Scheie, Scheie)	
MPS II (Hunteri sündroom)	
MPS IVA (Morquio A)	
MPS VI (Maroteaux–Lamy)	
MPS VII	
Pompe tõbi	
Wolmani tõbi	
Neuronaalne tseroidlipofustsinoos tüüp 2	
Niemann–Pick'i tõbi tüüp C	
MPS IIIA (Sanfilippo A)	Arendamisel
MPS IIIB (Sanfilippo B)	
Metakromaatiline leukodüstroofia	
Niemann–Pick'i tõbi tüüp B	
a-Mannosidoos	
Farberi tõbi	

(Beck, 2018; <https://mpssociety.org/learn/diseases/mps-vii/>)

MPS – mukopolüsahharidoos

1.5. Haigusseoselisuse hindamine

Kuigi NGS-andmete loomine on muutunud ajapikku odavamaks, on peamiseks tõkkeks nende andmete kasutamisel genotüübi-fenotüübi seoste tõlgendamine. Haigust põhjustavate variantide identifitseerimine tuhandete geneetiliste täisgenoomi (WGS, *whole-genome sequencing*) või täiseksoomi (WES, *whole-exome sequencing*) variantide seast nõuab mitmeid samme, milleks on variantide annoteerimine, filtreerimine, efekti *in silico* kahjulikkuse ennustamine ja kliinilise olulisuse tõlgendamine ekspertide poolt.

2015. aastal avaldasid ACMG ja Molekulaarpatoloogia Assotsiatsioon (AMP, *Association for Molecular Pathology*) suunised mendeliaarsete haigustega seotud geenivariantide tõlgendamiseks (Richards jt., 2015). Variantide kirjeldamiseks kasutatatakse viieastmelist klassifitseerimissüsteemi – patogeensed, tõenäoliselt patogeensed, ebaselge tähendusega, tõenäoliselt healoomulised ja healoomulised variandid. Tõenäoliselt patogeenne või tõenäoliselt healoomuline tähendab, et variandil on üle 90% tõenäosus olla vastavalt haigust põhjustav või healoomuline (Richards jt., 2015). Süsteem koosneb 28 kriteeriumist, mis põhinevad erinevatel andmetel, nt rahvastikuandmed, *in silico* andmed, funktsionaalsuse- ja segregatsioonandmed. Juhendi järgi viitavad patogeensusele näiteks null (-/-) variant geenis, milles funktsioonikadu on haiguse mehhanismiks; teadaoleva patogeense variandiga võrreldes sama aminohappe muutus olenemata nukleotiidi muutusest; *de novo* variant haigusega patsiendil; kui funktsionaalsed uuringud on näidanud kahjustavat efekti geenile ja geeniproduktile jne. Variandi healoomulisusele viitavad aga alleelisagedus üle 5% populatsiooni referentsandmebaasides või kui alleelisagedus on kõrgem kui haiguse puhul oodatav; kui variant, mille täielik penetrantsus peaks esinema varases eas, esineb tervel täiskasvanul; funktsionaalsed uuringud ei näita variandi kahjustavat efekti jne. Teabe kogumine iga kriteeriumi kohta võib olla aga üsna keeruline ja interpreteerijad võivad tõlgendada andmeid erinevalt (Li ja Wang, 2017).

Tõlgendamise abistamiseks on välja töötatud mitmeid tööriistu geneetiliste variantide funktsionaalse tähtsuse mõistmiseks seoses võimaliku mõjuga haigustele. Olemas on mitmed annotatsioonide genereerimise tööriistad, nt ANNOVAR (Yang ja Wang, 2015), dbNSFP (Liu jt., 2015) ja VEP (*Variant Effect Predictor*) (McLaren jt., 2016), mis näitavad, kuidas geenivariandid mõjutavad transkriptsiooni struktuuri või kodeerivat järjestust. Tööriistad klassifitseerivad variante introonseteks, intergeenseteks, splaiss-saidi ja eksoni variantideks ning viimaste jaoks näitavad, kuidas eksoni variandid mõjutavad aminohappejärjestust (Li ja

Wang, 2017). Kodeerivate variantide jaoks on olemas *in silico* arvutuslikud meetodid, näiteks SIFT (Kumar jt., 2009), PolyPhen-2 (Adzhubei jt., 2010), CADD (Kircher jt., 2014) ja MutationTaster (Schwarz jt., 2010), millega saab ennustada, kas variant on valkude funktsiooni või struktuuri kahjustav, kasutades evolutsioonilist informatsiooni, paiknemist valgujärjestuses ja biokeemilisi omadusi (Li ja Wang, 2017). Kuid neil on ka teatud piirangud, näiteks madal spetsiifilisus ja kahjulikkuse üleproгноosimine (Richards jt., 2015). Olemas on geeni- ja haiguspõhised andmebaasid (peatükk 1.5.1), kus dokumenteeritakse geenivariante, mis on kindlate haiguste suhtes patogeensed. Enamasti kogutakse sinna andmeid uuringute käigus või teaduslikust kirjandusest. Kuna erinevatel andmete sisestajatel on patogeensuse hindamiseks erinevad kriteeriumid, võib nende andmebaaside kirjade kvaliteet olla väga varieeruv ning variandid võivad olla klassifitseeritud valesti. Palju uuritud geenidel on kliinilistes andmebaasides rohkem sissekandeid ja nende interpretatsioon on tõenäoliselt täpsem (Li ja Wang, 2017).

Enne, kui indiviide teavitada patogeensest geenivariandist, tuleb kasutada interpretatsiooniks mitut allikat, sh kahjulikkuse ennustamise algoritme kuid üle vaadata ka saadavalolev kirjandus ja andmebaasid, et tagada patogeensuse hinnangule piisav toetus. Kuna andmebaaside hinnangud varieeruvad ning automatiseeritud algoritmid ei suuda tuvastada kõiki patogeensuse kriteeriumitele vastavaid variante, vajab geenivariantide kliinilise olulisuse hindamine manuaalset ülevaatamist. Geneetiline teave uueneb pidevalt ning geenivariantide hinnang võib aja jooksul ka muutuda (Blackburn jt., 2015).

1.5.1. Kliinilised andmebaasid

1.5.1.1. HGMD

HGMD (*Human Gene Mutation Database*) (Stenson jt., 2017) on ulatuslik geneetilisi haigusi põhjustavate pärilike mutatsioonide andmebaas, mis sisaldab üle 203 000 erineva geenivariandi (2017. a seisuga). HGMD-d kasutavad teadlased, arstid, diagnostikalaborid, molekulaargeneetikud, molekulaarbioloogid ja geneetilised nõustajad, aga ka biofarmatseudid ja bioinformaatikud üle maailma. Kõik HGMD mutatsioonide andmed on saadud teaduslikust kirjandusest ja neid kureeritakse käsitsi ekspertpaneelide poolt. HGMD-l on avalik versioon (<http://www.hgmd.org>), mis on kättesaadav akadeemiliste institutsioonide ja mittetulundusühingute registreeritud kasutajatele ning tasuline versioon (HGMD *Professional*), millele pääseb ligi litsentsi alusel. HGMD avalik versioon on tasulisest

versioonist umbes 3 aastat maas, sest andmeid uuendatakse avalikus versioonis ainult 2 korda aastas. Tasulises versioonis uueneb info pidevalt.

1.5.1.2. ClinVar

ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>; Landrum jt., 2016) on vabalt kättesaadav andmebaas, kus on tõlgendused variantide kliinilisest olulisusest.

Andmebaas on ulatuslik ning sisaldab erineva suuruse, tüübi või genoomse asukohaga iduliini ja somaatilisi variante. ClinVari andmestik on mõeldud selleks, et toetada näiteks meditsiinitöötajaid, kes soovib kiirelt kindlaks teha variandi tõlgenduse, milliseid haigusi on variandiga seoses kirjeldatud ja milline on variandi sagedus suuremahulistes populatsiooniuuringutes. ClinVar-i sisestavad interpretatsioone kliinilise testimise laborid, teaduslaborid, ekspertpaneelid ning infot saadakse uuringutest, teadustöödest, praktilistest juhistest ja teistest andmebaasidest. Andmete esitamisprotsessi käigus vaatavad ClinVar töötajad läbi HGVS (*Human Genome Variation Society*) nimetused (HGVS nomenklatuur on üle-maailmselt tunnustatud nomenklatuur geenivariantide kirjeldamiseks (den Dunnen jt., 2016)), haiguste nimetused, geeni-haiguse suhted ja andmebaasi identifikaatorid, kuid ei vaata üle kliinilist olulisust ega lahenda konflikte kliinilise olulisuse tõlgendamisel.

1.5.1.3. InterVar

InterVar (Li ja Wang, 2017) on vahend, mis aitab tõlgendada geenivariantide kliinilist olulisust ning põhineb 2015. aastal ACMG ja AMP poolt välja antud geenivariantide kliinilise tõlgendamise kriteeriumitel (peatükk 1.5). InterVar on mõeldud kasutamiseks teadlastele ja arstidele ning see aitab oluliselt kaasa haigusi põhjustavate geneetiliste variantide funktsionaalsete tagajärgede mõistmisele. InterVar on käsurealt juhitav tarkvara, millesse saab sisestada annotateeritud faili. Geenivariantide tõlgendamine toimub automaatselt, kuid kriteeriumeid on võimalik lisaks käsitsi kohandada. Variandid määratakse ACMG-AMP kriteeriumite kohaselt patogeenseks, tõenäoliselt patogeenseks, ebakindla tähendusega variandiks, tõenäoliselt healoomuliseks või healoomuliseks. InterVar on mõeldud ainult mendeliaarseid haigusi põhjustavate geneetiliste variantide tõlgendamiseks.

InterVar-il on ka veebiserver wInterVar (<http://wintervar.wglab.org/>), kuhu kasutajad saavad sisestada kromosomaalse positsiooni, geeni nime ja nukleotiidi muutuse või dbSNP-i identifikaatori. Samuti saab wInterVar-is lisaks automaatsele tõlgendamisele kriteeriume käsitsi reguleerida. wInterVar võimaldab aga tõlgendada ainult ühenukleotiidseid asendusi ning teiste variantide (nt indelite) interpreteerimiseks tuleb kasutada InterVar tarkvara.

1.5.1.4. Teised andmebaasid

Olemas on mitmeid teisi andmebaase, milles registreeritakse haigusega seotud variante.

OMIM (*Online Mendelian Inheritance in Man*) (<http://www.omim.org>; Amberger jt., 2015) on andmebaas, mis sisaldab ülevaateid geenidest ja geneetilistest haigustest. Informatsioon on kureeritud ekspertide poolt ning andmed saadakse biomeditsiinilisest kirjandusest.

dbSNP on andmebaas, mille eesmärgiks on ühte kohta kokku koguda identifitseeritud geneetilised variandid (ühenukleotiidsed variandid, indelid, lühikesed tandemkordused) ning sisaldab nii neutraalseid polümorfisme kui haigust põhjustavaid kliinilisi mutatsioone (<http://www.ncbi.nlm.nih.gov/SNP/>; Sherry jt., 2001).

LOVD (*Leiden Open Variation Database*) on süsteem, mis ühendab lookuspõhiste andmebaaside infot ning sisaldab fenotüüpidega seotud geenivariante ja informatsiooni indiviidide kohta, kellel variant leiti (<http://www.lovd.nl>; Fokkema jt., 2011).

2. EKSPERIMENTAALOSA

2.1. Töö eesmärgid

Käesoleva töö eesmärkideks on

1. Tuvastada Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) geenidoonoritel kliiniliselt olulisi (patogeenseid või tõenäoliselt patogeenseid) sekkumist võimaldavaid geenivariante konsensusnimekirjades esinevas 78 geenis (Tabel 1);
2. Võrrelda vastavate haiguste eeldatavaid sagedusi kirjanduses toodud andmetega;
3. Hinnata *in silico* patogeensust ennustavate skooride hinnangute kokkulangevusi teadaolevate patogeensuse hinnangutega;
4. Kaardistada lüsoosomaalsete ladestushaigustega seostatavate geenivariantide levimust Eestis, mille põhjal:
 - a. hinnata nendega seotud haiguste eeldatavat sagedust;
 - b. otsida potentsiaalseid kõrge haigusriskiga isikuid.

Lüsoosomaalsed ladestushaigused võimaldavad ensüümasendusravi kättesaadavuse pideva paranemise tõttu samuti sekkumist ning on seega potentsiaalselt muutunud efektiivset sekkumist võimaldavaks. Samuti on antud haigustel viimasel ajal hakatud leidma hilise avaldumisega vorme, mida võib esineda ka Geenivaramu valimis. Antud alalõigu puhul on tegemist Tartu Ülikooli ja Sanofi Aventis Eesti osakonna vahelise teaduskoostöö pilootprojektiga.

2.2. Materjal ja metoodika

2.2.1. Valim

Käesolevas töös kasutati TÜ Eesti Geenivaramu geenidoonorite 2420 täisgenoomi (WGS, *Whole Genome Sequencing*) ning 2356 täiseksoomi (WES, *Whole Exome Sequencing*) sekveneerimisel saadud andmeid. Kokku kasutati 4776 geenidoonori andmeid (edaspidi EGV NGS4776 andmestik), kellest 2268 on naised ning 2508 on mehed. Geenidoonorid on vanuses alates 18-ndast eluaastast.

2.2.2. Sekveneerimisandmed

Täisgenoomi sekveneerimine teostati Broadi Instituudis Cambridge'is USA-s, täiseksoomid sekveneeriti Nestlé Terviseteaduste Instituudis Lausanne'is. WGS proovid valmistati ette PCR-vaba meetodiga ning sekveneeriti Illumina HiSeq X Ten sekvenaatoriga, kasutades 150 bp paarisotsalisi lugemeid (*paired-end reads*) 30× keskmise katvusega. WES proovid sekveneeriti kasutades Agilent SureSelect Human All Exon V5+UTRs Kit'i vastavalt tootja soovitudele 67× keskmise *targeti* katvusega. Sekveneeritud lugemid joondati vastu GRCh37/hg19 inimese referentsgenoomi kasutades BWA-MEM (Li ja Durbin, 2009) v0.7.7 ning sekveneerimisel identifitseeritud variandid annoteeriti *Variant Effect Predictor* (VEP) (McLaren jt., 2016) versiooniga 88 ja ANNOVAR-iga (Yang ja Wang, 2015). Annotatsiooni valmimiseks vajalikud bioinformaatilised protseduurid olid eelnevalt teostatud TÜ doktorant Mart Kalsi poolt.

2.2.3. Variantide prioritiseerimine

Tööd alustati EGV NGS4776 annoteeritud andmetega. Andmetest selekteeriti patogeenselt või tõenäoliselt patogeenseid variante, mille puhul:

- on alleelisagedus kohordis <0,01;
- alleelisagedus <0,05 referentsandmebaasides gnomAD, 1000G ja ESP6500;
- *in silico* skoorid toetavad patogeensust;
- ClinVar andmebaasi hinnang patogeenne või tõenäoliselt patogeenne;
- InterVar andmebaasi hinnang patogeenne või tõenäoliselt patogeenne.

2.2.4. Alleelisageduse referentsandmebaasid

Et hinnata, kas geenivariante on eelnevalt täheldatud suurtes populatsioonipõhistes kohortides, on olemas mitmeid võrdlusandmebaase. Antud töös kasutati 3 andmebaasi andmeid, mis olid olemas annotatsioonifailides.

gnomAD (*The Genome Aggregation Database*) – sisaldab 123 136 eksoomi ja 15 496 täisgenoomi sekveneerimisandmeid, mis on saadud mitmetest suuremahulistest

populatsioonigeneetilistest ja haiguspõhistest uuringutest. Andmetest on eraldatud lapseas esinevate raskete haigustega indiviidid (mõned raskekujulise haigusega isikud võivad siiski andmetes esineda) ning nende esimese astme sugulased, seega on andmestikku võimalik kasutada alleelisageduse referentsina haruldaste haiguste uuringutes (<http://gnomad.broadinstitute.org/>).

1000G (*1000 Genomes Project*) – sisaldab 2650 indiviidi genotüüpe 26st populatsioonist Ameerikas, Aafrikas, Euroopas, Ida- ja Lõuna-Aasias, sisaldades kombineeritult madala katvusega täisgenoomi järjestust, kõrge katvusega eksoomi järjestust ja mikrokiibi genotüpiseerimist. Kõige uuemad andmed pärinevad faasist 3 (Auton jt., 2015).

ESP (*Exome Sequencing Project*) – eesmärgiks on tuvastada geneetilisi variante inimese genoomi valku kodeerivates regioonides. Kõige hilisemad ESP6500 andmed pärinevad 2203 Aafrika-Ameeriklaselt ning 4300 Euroopa-Ameeriklaselt, kokku 6503 indiviidilt (<http://evs.gs.washington.edu/EVS/>).

Töös välja toodud variantide alleelisagedus referentsandmebaasides on $<0,05$.

2.2.5. *In silico* skoorid

Käesolevas töös kasutati geenivariantide kahjuliku mõju hindamiseks kaheksat erinevat *in silico* ennustusliku meetodi skoori. Skooride väärtused olid olemas annotatsioonifailis ning nende interpretatsioon on toodud tabelis 3. Kasutusel olid CADD_Phred score (<http://cadd.gs.washington.edu>; Kircher jt., 2014), SIFT (<http://sift.bii.a-star.edu.sg>; Kumar jt., 2009), MutationAssessor (<http://mutationassessor.org>; Reva jt., 2011), PolyPhen-2 HumDiv ja HumVar (<http://genetics.bwh.harvard.edu/pph2>; Adzhubei jt., 2010), MutationTaster (<http://www.mutationtaster.org>; Schwarz jt., 2010), phyloP (Pollard jt., 2010) ning Condel (<http://bbglab.irbbarcelona.org/fannsdbs>; González-Pérez ja López-Bigas, 2011). Skoorid (v.a. CADD) on olemas nt ANNOVARis, kus neid saab annotatsiooniga integreerida. Kõigi 8 skooriga saab hinnata missenss-mutatsioone, kuid ainult 2 (CADD ja MutationTaster) võimaldavad hinnata ka nonsenss-, raaminihke ja splaiss-sait variantide patogeensust.

Tabel 3. *In silico* skoorid variantide patogeensuse hindamiseks

Skoori nimi	Ennustuse interpretatsioon
CADD_Phred (CADD)	Kõrgemad skoorid on kahjustavamad. Antud töös rakendati lävendina CADD>23; nonsenss- ja raaminihkemuutuse korral CADD>28.
SIFT (sift)	D: kahjustav (sift≤0.05); T: tolereeritud (sift>0.05)
PolyPhen 2 HDIV (pp2_hdiv)	D: tõenäoliselt kahjustav (≥0.957), P: võib-olla kahjustav (0.453≤pp2_hdiv≤0.956); B: healoomuline (pp2_hdiv≤0.452)
PolyPhen 2 HVar (pp2_hvar)	D: tõenäoliselt kahjustav (≥0.909), P: võib-olla kahjustav (0.447≤pp2_hdiv≤0.909); B: healoomuline (pp2_hdiv≤0.446)
MutationTaster (mt)	A" ("haigust_põhjustav_automaatne"); "D" ("haigust_põhjustav"); "N" ("polümorfism"); "P" ("polümorfism_automaatne")
MutationAssessor (ma)	H: kõrge; M: keskmine; L: madal; N: neutraalne. H/M on funktsionaalne ja L/N mittefunktsionaalne.
phyloP (phylop)	Kõrgemad skoorid on kahjustavamad. Antud töös rakendati lävendina phylop>3
Condel	D: kahjustav; N: neutraalne

(<http://annovar.openbioinformatics.org/en/latest/user-guide/filter/>, muudetud tabel)

2.2.6. Kasutatud andmebaasid

Töös lähtuti tulemuste väljatoomisel lisaks ka kliiniliste andmebaaside hinnangutest. Kasutati kahte andmebaasi – ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) ja veebipõhine InterVar (<http://wintervar.wglab.org/>). ClinVar variantide esialgne hinnang oli annotatsioonifailis olemas, kuid töö käigus kontrolliti andmed käsitsi üle, kuna *online* andmebaase uuendatakse esimesena. Andmebaaside otsingu aluseks oli variandi rs number.

InterVar andmebaasi hinnangut annotatsioonifailis ei olnud ning see leiti käsitsi. Andmebaasi sisestati variandi rs number, selle puudumisel kromosoomi, positsiooni ja nukleotiidi muutuse info. Veebipõhine InterVar andmebaas võimaldab interpretatsioone ainult eksonipõhistele ühenukleotiidsetele variantidele.

Töö tulemuste esitamisel lähtuti sellest, et geenivariant oleks vähemalt ühe andmebaasi põhjal patogeenne või tõenäoliselt patogeenne.

2.2.7. Arvutused

Töös arvutati vastavalt saadud tulemustele haiguste eeldatav esinemise sagedus Eestis. Arvutused põhinevad Hardy-Weinbergi printsiibil. Autosoom-retsessiivsete haiguste puhul kasutati arvutuste toetuseks veebipõhist alleelisageduste kalkulaatorit (<https://wpcalc.com/en/equilibrium-hardy-weinberg/>). X-liitelise haiguse korral arvutati sagedus naiste-meeste puhul eraldi. Arvutustest eemaldati I astme sugulased.

2.2.8. Lüsosomaalsete ladestushaigustega seotud geenid

Töö teisest eesmärgist lähtuvalt valiti töösse geenid, millega seotud lüsosomaalsetest ladestushaigustest enamusele on ravi olemas või arendamisel (Tabel 2). Geenid ja vastavad haigused on toodud tabelis 4.

Tabel 4. Geenid ja vastavad lüsosomaalsed ladestushaigused

Geen	Haigus	MIM #	Pärandumine
<i>GAA</i>	Pompe tõbi	232300	AR
<i>GLA</i>	Fabry tõbi	301500	XL
<i>GBA</i>	Gaucher' tõbi	230800; 230900; 231000; 231005	AR
<i>SMPD1</i>	Niemann-Pick'i tõbi tüüp A, tüüp B	257200; 607616	AR
<i>NPC1</i>	Niemann-Pick'i tõbi tüüp C1	257220	AR
<i>NPC2</i>	Niemann-Pick'i tõbi tüüp C2	607625	AR
<i>IDUA</i>	MPS IH e. Hurleri sündroom MPS IH/S e. Hurler-Scheie sündroom MPS IS e. Scheie sündroom	607014; 607015; 607016	AR
<i>IDS</i>	MPS II e. Hunteri sündroom	309900	XLR
<i>SGSH</i>	MPS IIIA e. Sanfilippo sündroom tüüp A	252900	AR
<i>NAGLU</i>	MPS IIIB e. Sanfilippo sündroom tüüp B	252920	AR
<i>HGSNAT</i>	MPS IIIC e. Sanfilippo sündroom tüüp C	252930	AR
<i>GNS</i>	MPS IIID e. Sanfilippo sündroom tüüp D	252940	AR
<i>GALNS</i>	MPS IVA e. Morquio sündroom tüüp A	253000	AR
<i>GLB1</i>	MPS IVB e. Morquio sündroom tüüp B	253010	AR
<i>ARSB</i>	MPS VI e. Maroteaux-Lamy sündroom	253200	AR
<i>GUSB</i>	MPS VII e. Sly sündroom	253220	AR

AD – autosoom-dominantne, AR – autosoom-retsessiivne, XL – X-liiteline; XLR – X-liiteline retsessiivne; MIM # – haiguse identifikaator OMIM andmebaasis

2.3. Tulemused ja arutelu

2.3.1. ACMG/Geisinger nimekirjaga seotud tulemused

Töö käigus analüüsiti ACMG/Geisingeri nimekirja 78 geenis ühenukleotiidsetest asendustest, väikestest insertioonidest ja deletsioonidest põhjustatud raaminihkkeid, missenss-, nonsenss- ja (kanoonilisi) splaiss-sait mutatsioone, mis võivad põhjustada geeni funktsiooni kadu või selle häirumist. Töös tuvastati patogeenseid geenivariante *in silico* skooride kaudu, alleelisageduse järgi kohordis ja referentsandmebaasides ning tuginedes kliiniliste andmebaaside infole. Töös toodi välja geenivariandid, millel on olemas andmebaasides patogeensust toetav hinnang. Kokku leiti 77 patogeenset või tõenäoliselt patogeenset geenivarianti 36 geenis (Tabel 5). Tabelis on välja toodud lisaks geenile ja variandile dbSNP rs number ning selle puudumisel kromosoom ja positsioon, ClinVar ja InterVar andmebaaside hinnang, isikute arv, kellel vastav variant esineb ning *in silico* skooride arv, mis ennustasid variandi patogeenseks. Geenidele vastavad haigused on välja toodud töö kirjanduse osas (Tabel 1). Patogeensed geenivariandid esinevad kokku 237 indiviidil, kellest 4 kannavad korraga kahte erinevat varianti kahes eri geenis (Tabel 6).

Tabel 5. Geenivariandid vastavalt ACMG/Geisingeri nimekirja geenidele

Geen	Variant	dbSNP ID või chr:pos	Clin-Var	Inter-Var	Isikute arv (n=4776)	Patogeensuse skoorid
<i>ACVRL1</i>	c.920C>T p.Ala307Val	12:52309156	.	TP	1	4/8
<i>APC</i>	c.481C>T p.Gln161*	rs876658325	P	P	1	2/2
<i>APOB</i>	c.10898G>A p.Trp3633*	2:21228842	.	P	1	2/2
<i>ATP7B</i>	c.4135C>T p.Pro1379Ser	rs181250704	VUS	TP	3	6/8
	c.3895C>G p.Leu1299Val	13:52511620	.	TP	1	8/8
	c.3818C>T p.Pro1273Leu	rs758355520	P/TP	TP	1	8/8
	c.3505A>G p.Met1169Val	rs749085322	TP	TP	1	7/8
	c.3207C>A p.His1069Gln	rs76151636	P	TP	64	7/8
	c.2605G>A p.Gly869Arg	rs191312027	P/TP	TP	1	8/8
	c.2532delA p.Val845Serfs*28	rs755709270	P	.	1	2/2
	c.1630C>T p.Gln544*	rs766906034	.	P	4	2/2
	c.658G>T p.Gly220*	13:52548698	.	P	1	2/2
<i>BRCA1</i>	c.5329_5330insC p.Gln1777Profs*74	rs397507247	P	.	8	1/2

	c.4258C>T p.Gln1420*	rs80357305	P	P	1	2/2
	c.4065_4068delTCAA p.Asn1355Lysfs*10	rs80357508	P	.	1	1/2
	c.4035delA p.Glu1346Lysfs*20	rs80357711	P	.	11	0/2
	c.1840A>T p.Lys614*	rs80357282	P	P	1	2/2
	c.1687C>T p.Gln563*	rs80356898	P	P	1	2/2
<i>BRCA2</i>	c.9381G>A p.Trp3127*	rs876661242	P	P	1	2/2
	c.37G>T p.Glu13*	rs80358622	P	P	1	2/2
	c.6402_6406delTAACT p.Asn2135Lysfs*3	rs80359584	P	.	1	1/2
	c.8572C>T p.Gln2858*	rs80359112	P	P	4	2/2
<i>CACNA1S</i>	c.2707C>T p.Arg903*	rs367983954	VUS	P	1	2/2
	c.502C>T p.Arg168*	rs201998231	TP	P	2	2/2
<i>CAV3</i>	c.277G>A p.Ala93Thr	rs28936686	P/TP	VUS	2	8/8
<i>COL3A1</i>	c.1475G>A p.Gly492Glu	rs587779500	P	VUS	1	7/8
	c.1714C>T p.Arg572*	rs572097661	.	P	1	2/2
<i>DSG2</i>	c.2315T>G p.Leu772*	rs794728097	TP	P	1	2/2
<i>DSP</i>	c.4357C>T p.Gln1453*	6:7580780	.	P	1	2/2
<i>FBN1</i>	c.7339G>A p.Glu2447Lys	rs137854464	P/TP	TP	1	8/8
<i>GLA</i>	c.427G>A p.Ala143Thr	rs104894845	C	TP	2	7/8
<i>KCNE2</i>	c.79C>T p.Arg27Cys	rs74315449	C	TP	3	6/8
	c.161T>C p.Met54Thr	rs74315447	P/TP	TP	1	5/8
<i>KCNH2</i>	c.2587C>T p.Arg863*	rs773724817	P	P	1	2/2
<i>KCNQ1</i>	c.477+1G>A	rs762814879	P	.	1	2/2
	c.604G>A p.Asp202Asn	rs199472702	P	P	1	8/8
	c.1664G>A p.Arg555His	rs199472800	P	TP	1	8/8
	c.683+5G>A	rs397508122	P	.	1	0/2
<i>LDLR</i>	c.643C>T p.Arg215Cys	rs764042910	P	TP	1	7/8
	c.768C>A p.Asp256Glu	19:11217314	.	TP	1	7/8
	c.1202T>A p.Leu401His	rs121908038	P/TP	VUS	2	7/8
	c.1335C>A p.Asp445Glu	rs749780672	TP	VUS	1	7/8
<i>MLH1</i>	c.1668-1G>T	rs267607845	TP	.	1	2/2
<i>MSH2</i>	c.2131C>T	rs63750636	P	P	1	2/2

	p.Arg711*					
<i>MSH6</i>	c.1095G>A p.Trp365*	2:48026217	.	P	1	2/2
	c.3226C>T p.Arg1076Cys	rs63750617	TP	TP	1	5/8
	c.742C>G p.Arg248Gly	rs63749980	VUS	TP	1	5/8
<i>MUTYH</i>	c.725G>A p.Arg242His	rs140342925	P/TP	TP	1	8/8
	c.316C>T p.Arg106Trp	rs765123255	P/TP	TP	1	7/8
	c.527A>G p.Tyr176Cys	rs34612342	P	P	13	8/8
	c.1178G>A p.Gly393Asp	rs36053993	P/TP	P	24	8/8
<i>MYH7</i>	c.2348G>A p.Arg783His	rs397516142	C	TP	1	2/8
	c.5655G>A p.Ala1885=	rs753392652	P	VUS	1	0/2
<i>MYL2</i>	c.401A>C p.Glu134Ala	rs143139258	C	TP	1	6/8
<i>MYL3</i>	c.461G>A p.Arg154His	rs104893749	C	TP	1	6/8
	c.262A>G p.Thr88Ala	3:46902211	.	TP	1	7/8
	c.254A>G p.Gln85Arg	3:46902219	.	TP	1	6/8
	c.170C>A p.Ala57Asp	rs139794067	C	TP	1	8/8
<i>PMS2</i>	c.861_864delACAG p.Arg287Serfs*19	rs267608154	P	.	1	1/2
<i>RET</i>	c.2410G>A p.Val804Met	rs79658334	P/TP	P	2	7/8
<i>RYR1</i>	c.325C>T p.Arg109Trp	rs118192173	P/TP	TP	3	7/8
	c.1840C>T p.Arg614Cys	rs118192172	P/TP	TP	10	7/8
	c.7268T>A p.Met2423Lys	rs118192174	P	VUS	1	7/8
	c.9579C>G p.Cys3193Trp	rs587784379	P	VUS	4	6/8
	c.14385G>A p.Trp4795*	rs565173276	.	P	1	2/2
<i>RYR2</i>	c.3151C>T p.Arg1051Cys	rs533330664	C	TP	1	4/8
<i>PKP2</i>	c.2146-1G>C	rs193922674	P/TP	.	1	2/2
<i>SCN3B</i>	c.29T>C p.Leu10Pro	rs121918282	C	TP	7	1/8
<i>SCN5A</i>	c.5872C>T p.Arg1958*	rs757532106	TP	VUS	1	2/2
	c.3691G>A p.Glu1231Lys	rs199473598	P	TP	1	2/8
	c.2911C>T p.Arg971Cys	rs61737825	C	TP	2	3/8
	c.655C>T p.Arg219*	rs577421914	.	P	2	2/2
<i>SDHC</i>	c.405+1G>T	rs587776653	P	.	1	1/2
<i>STK11</i>	c.368A>G p.Gln123Arg	rs764449808	TP	TP	1	5/8
<i>TNNI3</i>	c.422G>A	rs397516347	P/TP	VUS	1	6/8

	p.Arg141Gln					
<i>TNNT2</i>	c.780+5G>A	rs730881113	TP	.	13	0/2
<i>TP53</i>	c.542G>A p.Arg181His	rs397514495	P/TP	TP	1	7/8

P – patogeenne

TP – tõenäoliselt patogeenne

C – (*conflicting interpretation*) patogeensuse hinnangutes esineb konflikte (P, TP, VUS)

VUS – (*variant of uncertain significance*) ebaselge tähendusega variant

“.” – infot ei ole

Tulemuste välja toomisel lähtuti sellest, et geenivariandi patogeensusel oleks ka kliiniliste andmebaaside toetav hinnang. Töös kasutati ClinVar ja InterVar andmebaase. Peaaegu kolmandikul juhtudest (n=28) on variantidel olemas mõlema andmebaasi patogeensust toetav hinnang, ainult ClinVar (n=24) või ainult InterVar (n=25) hinnang.

Variantide patogeensuse hindamisel tugineti *in silico* ennustuslikele skooridele. Hindamisel kasutati 8 skoori, mille kõigiga saab hinnata missenss-tüüpi mutatsioone, kuid ainult 2 (CADD ja MutationTaster) on mõeldud ka nonsenss-, raaminihke ja splaiss-sait variantide patogeensuse hindamiseks. Skooride väärtused variantide kaupa on toodud töö lisas (Lisa 1). Tulemustest selgus, et 36 juhul 77st ennustasid variandi patogeensust kõik skoorid ning 33 juhul ennustasid patogeensust vähemalt pooled skoorid. 4 juhul ennustasid patogeensust alla poole skooridest ning 4 juhul ei ennustanud patogeensust üksi skoor.

Näiteks ei ennustanud skoorid patogeensust *BRCA1* c.4035delA (p.Glu1346Lysfs*20) puhul. On aga teada, et antud variant on üks enam levinumatest rinna- ja munasarjavähi eelsoodumust tekitavatest geenivariantidest Euroopas, sealhulgas Eestis (Tamboom jt., 2010). Splaiss-sait mutatsiooni *KCNQ1* c.683+5G>A ning sünonüümse asenduse *MYH7* c.5655G>A (p.Ala1885=) puhul ei ennustanud patogeensust samuti ükski skoor, kuid kirjanduse andmetel põhjustavad mõlemad mutatsioonid eksoni vahelejätmist ning kõrvalekaldeid splaissingul (Crehalet jt., 2012; Pajusalu jt., 2016). *KCNQ1* c.683+5G>A esineb ka HGMD andmebaasis seoses pika QT sündroomiga.

SCN3B c.29T>C (p.Leu10Pro) puhul ennustab variandile patogeensust ainult 1 skoor 8-st. Variant on esinenud Brugada sündroomiga patsiendil ning geeni ekspressiooni analüüsil leiti variandi põhjuslik seos haigusega (Hu jt., 2009), variant esineb ka seoses Brugada sündroomiga HGMD andmebaasis. *MYH7* c.2348G>A (p.Arg783His) puhul ennustavad kahjulikkust 2 skoori 8-st, variant on leitud kardiomiopaatia patsientidel (Waldmüller jt., 2008; Kostareva jt., 2016) ning esineb seoses hüpertroofse kardiomiopaatia ka HGMD andmebaasis.

Tabel 6. ACMG/GHS76 nimekirja topeltkandlused

Isik	Geen	Variant
Isik 1	<i>ATP7B</i>	c.3207C>A (p.His1069Gln)
	<i>BRCA1</i>	c.4035delA (p.Glu1346Lysfs*20)
Isik 2	<i>ATP7B</i>	c.3207C>A (p.His1069Gln)
	<i>MUTYH</i>	c.1178G>A (p.Gly393Asp)
Isik 3	<i>RYR1</i>	c.325C>T (p.Arg109Trp)
	<i>SCN5A</i>	c.655C>T (p.Arg219*)
Isik 4	<i>GLA</i>	c.427G>A (p.Ala143Thr)
	<i>MSH6</i>	c.742C>G (p.Arg248Gly)

Tabelis 6 on välja toodud isikud, kellel esineb korraga kaks erinevat antud töös leitud geenivarianti. Isik 1 kannab korraga nii Wilsoni tõve kui rinna- ja munasarjavähiga seotud geenivarianti. *ATP7B* p.His1069Gln mutatsioon on Euroopas sagedaseim Wilsoni tõbe põhjustav alleel, mis esineb keskmiselt 35-45% haigetel (<https://www.kliinikum.ee/geneetikakeskus/molekulaardiagnostika/48-wilsoni-tobi>), Wilsoni tõbi on autosoom-retsessiivne haigus ning isik 1 on alleeli suhtes heterosügootne, seega ei ole haigusest mõjutatud, kuid võib selle järglastele edasi pärandada. Kuna *BRCA1* c.4035delA (p.Glu1346Lysfs*20) on üks enam levinumatest rinna- ja munasarjavähi eelsoodumust tekitavatest geenivariantidest, võib isikul 1 olla rinna- ja munasarjavähi kõrgem haigusrisk. Tegemist on 1975. aastal sündinud naisega, kellel terviseandmetes hetkel rinnavähi diagnoosi ei ole.

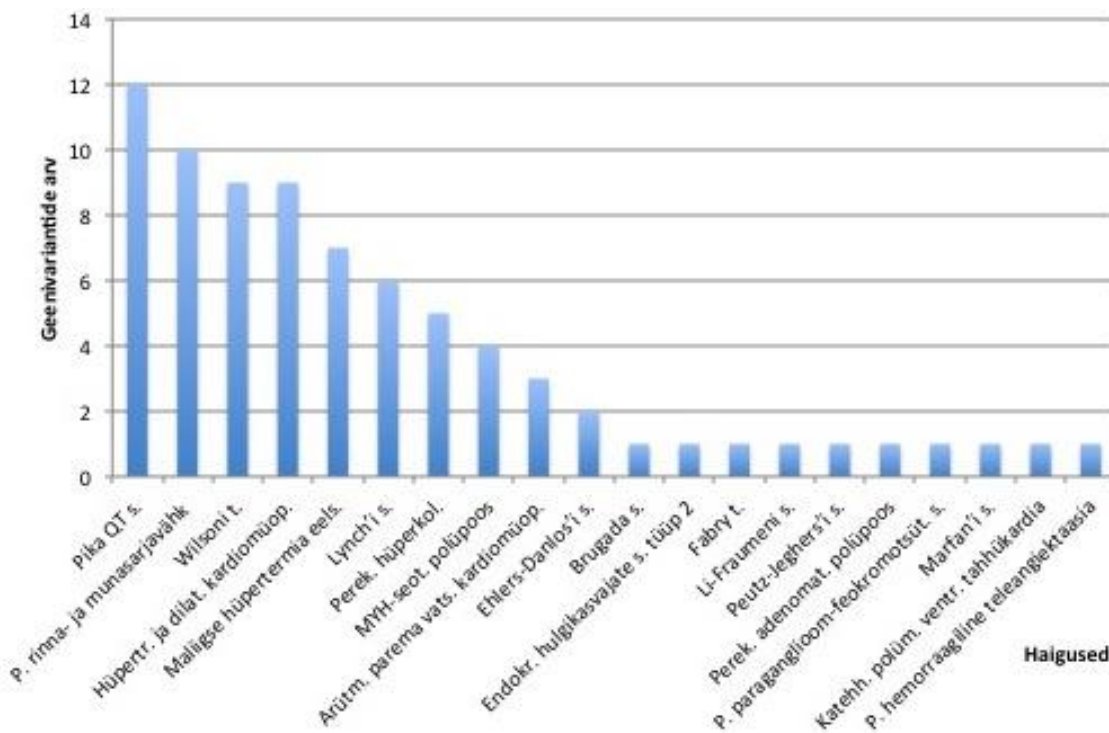
Isik 2 kannab lisaks Wilsoni tõve alleelile ka MYH-seotud polüpoosi alleeli. *MUTYH* c.1178G>A (p.Gly393Asp), mis on üks kõige tavalisematest MYH-seotud polüpoosi põhjustajatest, esinedes 50–82% patsientidest (Aretz jt., 2014). Antud haigus on samuti autosoom-retsessiivne ning isik alleeli suhtes heterosügootne.

Isikul 3 võib olla nii maliigse hüpertermia kui pika QT sündroomi kõrge haigusrisk, kuid diagnoosid hetkel haigustele ei viita.

Isik 4, kes on 1952. aastal sündinud naine, kannab X-liitelise Fabry tõve geenivarianti c.427G>A (p.Ala143Thr), kuid tal esineb ka Lynch'i sündroomi seoseline variant *MSH6* c.742C>G (p.Arg248Gly). Fabry tõvega seostatud geenivariant kuulub antud töös lisaks ACMG/GHS76 nimekirjale ka lüsoosomaalsete ladestushaiguste nimekirja ning seda käsitletakse pikemalt peatükis 2.3.2. ACMG ja Geisinger-76 nimekirjades on soovitatud tagasisidet anda Fabry tõve kardiomiopaatia vormile, kuid antud geenivariandi puhul ei ole teada, kas see võiks põhjustada kardiomiopaatiat. *MSH6* c.742C>G (p.Arg248Gly) on

ClinVar andmebaasis ebaselge tähendusega. 5 skoori 8-st hindavad variandile patogeensust ning InterVar hinnangul on variant tõenäoliselt patogeenne, seega võib tegemist olla kõrge haigusriskiga, kuid selget seost ei ole teadaolevalt kirjeldatud. Variant puudub ka HGMD andmebaasist. Isik 4 diagnooside andmed Lynchi sündroomile hetkel ei viita.

Käesoleva töö tulemustest selgus (Joonis 1), et kõige rohkem geenivariante (n=12) esines seoses pika QT sündroomiga geenides *CAV3*, *KCNQ1*, *KCNH2*, *KCNE2* ja *SCN5A*. Sellele järgnes geenivariantide arvu poolest pärilik rinna- ja munasarjavähk, mille puhul esines 10 varianti geenides *BRCA1* ja *BRCA2*. 9 geenivarianti leiti seoses Wilsoni tõvega geenis *ATP7B*. Samuti 9 varianti leiti seoses hüpertroofilise ja dilatatiivse kardiomiopaatia, geenides *MYH7*, *TNNT2*, *TNNI3*, *MYL3* ja *MYL2*. 7 geenivarianti leiti seoses maliigse hüpertermia eelsoodumusega geenides *RYR1* ja *CACNA1S*. 6 geenivarianti leiti Lynch'i sündroomi seoselistes geenides *MLH1*, *MSH2*, *MSH6*, *PMS2*. 5 geenivarianti perekondliku hüperkolesteroleemiaga seoselistes geenides *LDLR*, *APOB*. 4 *MYH*-seoselisele polüpoosile vastavat geenivarianti leiti geenis *MUTYH*. 3 geenivarianti leiti seoses arütmogeense parema vatsakese kardiomiopaatia geenides *PKP2*, *DSP*, *DSG2*. 2 geenivarianti leiti Ehlers-Danlos'i sündroomi seoselises geenis *COL3A1*. Brugada sündroomi (*SCN3B*), II tüüpi endokriinsete hulgikasvajate sündroomi (*RET*), Fabry tõve (*GLA*), Li-Fraumeni sündroomi (*TP53*), Peutz-Jeghers'i sündroomi (*STK11*), perekondliku adenomatoosse polüpoosi (*APC*), päriliku paraganglioom-feokromotsütoomi sündroomi (*SDHC*), Marfan'i sündroomi (*FBN1*), katehoolaminergilise polümorfse ventrikulaarse tahhükardia (*RYR2*) ja päriliku hemorraagilise teleangiektaasia (*ACVRL1*) seoselistes geenides leiti kõigis 1 geenivariant.



Joonis 1. Patogeensete või tõenäoliselt patogeensete geenivariantide arv vastavalt ACMG/GHS76 nimekirja haigustele.

Käesoleva töö tulemustest selgus (Joonis 2), et 77 indiviidil esineb Wilsoni tõve seoseline geenivariant ning 39 indiviidil MYH-seotud polüpoosi seoseline geenivariant. Kuna Wilsoni tõbi ning MYH-seotud polüpoos on autosoom-retsessiivselt päranduvad haigused, on tegemist kandjatega, sest seoses vastavate haigustega ühtegi alternatiivset homosügooti ega liitheterosügooti antud töös ei leitud. Ka Fabry tõve seoseliste geenivariantidega isikute puhul on tegemist kandjatega ($n=2$), sest Fabry tõbi on X-liitelise pärandumisega ning geenidoonorite andmed näitavad, et tegemist on heterosügootsete naistega.

Kõik teised haigused päranduvad autosoom-dominantselt. Seega potentsiaalselt kõrge haigusriskiga indiviididest enamusel ($n=30$) esineb päriliku rinna-ja munasarjavähi seoseline geenivariant. 22 indiviidil esineb maliigse hüpertermia eelsoodumusega seostatud geenivariant, 21 indiviidil hüpertroofilise/dilatatiivse kardiomiopaatiaga seostatud variant ning 17 indiviidil pika QT sündroomi seoseline geenivariant. Brugada sündroomiga seostatud geenivariant esineb 7 indiviidil, Lynch'i sündroomiga seostatud geenivariant 6 indiviidil ning perekondliku hüperkolesteroleemiaga seostatud geenivariant samuti 6 indiviidil. 3 indiviidil esineb arütmogeense parema vatsakese kardiomiopaatiaga seostatud geenivariant. 2 indiviidil esineb II tüüpi endokriinsete hulgakasvajate sündroomi seoseline variant ning samuti 2

Wilsoni tõbi ja MYH-seotud polüpoos on autosoom-retsessiivselt päranduvad haigused ning Fabry tõbi X-liiteliselt päranduv, seega peedeldab joonis antud haiguste puhul kandjate arvu (Fabry tõvega seotud isikud on naised), ülejäänud haiguste puhul potentsiaalselt kõrge haigusriskiga iskute arvu.

Tabel 7. ACMG/GHS76 nimekirja haiguste eeldatav sagedus Eestis võrdluses kirjanduses väljatoodud sagedusega

Haigus	Sagedus	Sagedus kirjanduses ¹
Wilsoni tõbi	1/15 200	1/30 000
MYH-seotud polüpoos	1/56 600	1/20 000 - 1/40 000
Pärilik rinna- ja munasarjavähk	1/160	1/400 - 1/500
Maliigse hüpertermia eelsoodumus	1/210	1/200 - 1/5000
Hüpertroofiline ja dilatatiivne kardiomüopaatia	1/230	1/250 - 1/500
Pika QT sündroom	1/270	1/2500
Brugada sündroom	1/660	1/2000
Lynch'i sündroom	1/770	1/440
Perekondlik hüperkolesteroleemia	1/770	1/200 - 1/250
Arütmogeenne parema vatsakese kardiomüopaatia	1/1540	1/1000 - 1/5000
Ehlers-Danlos'i sündroom, vaskulaarne tüüp	1/2300	1/10 000 - 1/25 000
Endokriinsete hulgikasvajate sündroom tüüp 2	1/2300	1/35 000
Fabry tõbi	1/1130*	1/3000 - 100 000
Li-Fraumeni sündroom	1/4600	1/5000 - 1/20 000
Peutz-Jeghers'i sündroom	1/4600	1/25 000
Perekondlik adenomatoosne polüpoos	1/4600	1/6800 - 1/31 200
Pärilik paraganglioom-feokromotsütoomi sündroom	1/4600	1/1700 - 1/4500
Marfan'i sündroom	1/4600	1/5000
Katehhoolaminergiline polümorfne ventrikulaarne tahhükardia	1/4600	1/2000 - 1/10 000
Pärilik hemorraagiline teleangiiektaasia	1/4600	1/5000 - 1/8000

¹Sageduse andmed põhinevad GeneReviews® (<https://www.ncbi.nlm.nih.gov/books/NBK11116/>); Orphanet (<https://www.orpha.net/>); (<https://www.uptodate.com/contents/susceptibility-to-malignant-hyperthermia-evaluation-and-management>); (Kantorovich jt., 2010). *Fabry tõve puhul on tegemist X-liitelise haigusega ning antud tulemuste põhjal on välja toodud sagedus naiste seas.

Rinna- ja munasarjavähi patogeense geenivariandi esinemisel võib elu jooksul rinnavähki haigestumise tõenäosuseks olla 46% kuni 87%, munasarjavähi puhul 16.5% to 63% (Petrucelli jt., 2016). Haiguse ennetamiseks tuleks teha iga-aastaseid sõeluuringuid, väga

kõrge riski korral kaaluda profülaktilist ooforektoomiat või mastektoomiat. Ka meestel võib avalduda rinnavähk või seoses *BRCA1/2* geenidega eesnäärmevähk, seega tuleks ka mehi patogeense leiu korral kontrollida.

Maliigse hüpertermia eelsoodumus on üheks võimalikuks surma põhjustajaks läbi anesteesi. Vältida tuleks teatud anesteetikume, samuti kõrgeid temperatuure. Pika QT sündroom võib põhjustada äkksurma ning beeta-blokaatoreid soovitatakse ka asümptomaatilistele isikutele, kellel on pika QT kõrge risk (Alders jt., 2018), samuti peaks vältima tugevat füüsilist aktiivsust. Hüpertroofiline kardiomiopaatia võib kaasuvate riskifaktorite esinemisel põhjustada südameseiskusest tingitud äkksurma. Soovitatakse implanteeritava kardioverter-defibrillaatori paigaldamist ja hoidumist võistlusspordist. Dilatatiivse kardiomiopaatia puhul tuleks haiguse ennetamiseks iga 1-3 aasta tagant läbi viia kardiovaskulaarne skriining (<https://www.ncbi.nlm.nih.gov/books/NBK1116/>).

Käesoleva töö käigus leiti ACMG/GHS76 nimekirjade alusel 77 patogeenset või tõenäoliselt patogeenset geenivarianti 36 geenis. Leitud geenivariandid esinevad kokku 237 indiviidil 4776-st, nendest 118 on seotud autosoom-retsessiivse või X-liitelise haigusega ning haiguste kandjad, seega potentsiaalseid kõrge haigusriskiga isikuid võiks olla 119. Geenivariantide patogeensust hinnati alleelisageduse, kliiniliste andmebaaside ja *in silico* skooride alusel. Erinevusi patogeensuse hindamisel esines andmebaaside ja skooride vahel vaid üksikjuhtudel. Töös leiti ka, et kõige rohkem võiks nimekirjas olevatest haigustest esineda Eestis pärilikku rinna- ja munasarjavähki, maliigse hüpertermia eelsoodumust, hüpertroofilist ja dilatatiivset kardiomiopaatiat ning pika QT sündroomi. Nendest pärilik rinna- ja munasarjavähk ning pika QT sündroom on antud tulemuste põhjal Eestis eeldatavalt sagedasemad kui üldises populatsioonis, maliigse hüpertermia eelsoodumus ning hüpertroofilise ja dilatatiivse kardiomiopaatia sagedused on aga võrreldes üldise populatsiooniga samad. Potentsiaalselt kõrge haigusriski korral tuleks kaaluda personaliseeritud tagasisidet, et võimaldada inimestele õigeaegset ravi ning vähendada haigusest tulenevaid komplikatsioone.

2.3.2. Lüsosomaalsete ladestushaigustega seotud tulemused

Töö käigus analüüsiti 16 lüsosomaalsete ladestushaigustega seotud geenis ühenukleotiidsetest asendustest, väikestest insertioonidest ja deletsioonidest põhjustatud raaminihkeid, missenss-, nonsenss- ja (kanoonilisi) splaiss-sait mutatsioone, mis võivad põhjustada geeni funktsiooni kadu või selle häirumist. Kokku leiti 48 patogeenset või tõenäoliselt patogeenset varianti 15 geenis (Tabel 8). Vastavaid variante leiti kokku 126 indiviidil, kellest 4 isikul on

korraga kaks erinevat varianti (Tabel 9). 1 *GLB1* geeni variantidest seostatakse GM1-gangliosidoosiga (GM1-g), mis on samuti lüsosomaalne ladestushaigus, kuid ei kuulu antud töös käsitletud haiguste nimekirja. Ühtegi haigusseoselist varianti ei leitud *IDS* geenis, mis on seotud MPS II-ga.

Kuna lüsosomaalsed ladestushaigused on retsessiivse pärandumismustriga, siis avaldub haigus alternatiiv-homosügootses, liitheterosügootses või hemisügootses olekus. Antud tulemustes leiti üks liitheterosügootne kandja, kellel on ühe geeni kaks erinevat varianti (Tabel 8). *GLA* geeni puhul on tegemist X-liitelise pärandumismustriga. On teada, et käesolevas töös on kõik *GLA* variandiga indiviidid naised.

Töös kasutati patogeensuse hindamiseks ClinVar ja InterVar andmebaase. Peaaegu pooltel juhtudel (n=22) on variantidel olemas mõlema andmebaasi patogeensust toetav hinnang, teistel juhtudel ainult ClinVar (n=15) või InterVar (n=11) hinnang.

Toetavalt vaadati *in silico* ennustavaid skooride. Tabelis 8 on välja toodud skooride arv, mis ennustavad variandile kahjulikkust. 8st kasutatud skoorist võimaldavad kõik hinnata missenss-muutusi, kuid ainult 2 skoori on mõeldud nonsenss- ja raaminihke ning splaiss-sait variantide patogeensuse hindamiseks. Skooride täpsemad andmed variantide kaupa on toodud töö lisas (Lisa 2). Tulemustest selgus, et 25 juhul 48st ennustasid variandile patogeensust kõik skoorid ning 16 juhul ennustasid patogeensust vähemalt pooled skoorid. 2 juhul ennustasid patogeensust alla poole võimalikest skooridest ning 5 juhul ei ennustanud patogeensust ükski skoor.

Ükski skoor ei toetanud patogeensust *GAA* c.1552-3C>G ja c.-32-13T>G, *GLB1* c.75+2dupT, *NPC2* c.190+5G>A ja *SMPD1* c.1826_1828delGCC variantidele. 4 esimest varianti on esindatud HGMD andmebaasis. *GAA* c.1552-3C>G variant muudab splaissingu akseptorsaiti ning eksonis paiknevat splaissingu võimendajat (ESE, *exonic splicing enhancer*). *NPC2* c.190+5G>A ja *GLB1* c.75+2dupT muudavad splaissingu doonorsaiti *SMPD1* c.1826_1828delGCC mõjutab ESE saiti ning tekitab eksonis paikneva splaissingu vaigistaja (ESS, *exonic splicing silencer*) saidi. Seega mõjutavad antud variandid splaissingut ning on leitud ka vastavate haigustega patsientidel (Kroos jt., 2006; Mavridou jt., 2016). *GLB1* c.75+2dupT mutatsiooni seostatakse GM1-gangliosidoosiga (Chakraborty jt., 1994), mida antud töö haiguste nimekirjas ei ole. *GAA* c.-32-13T>G varianti on samuti korduvalt leitud Pompe tõvega patsientidel ning põhjustab kergemat Pompe haigusvormi (Kroos jt., 2006). Antud juhtudel on tegemist splaiss-sait variantide ning raaminihet mittepõhjustava deletsiooniga, millel on hindamiseks vähem skoor kui missenss-variantidel ning on teada, et

skoorid ei ole sellistel juhtudel väga täpsed. Seega skooride puudumine ei tähenda alati, et variant ei ole patogeenne ning oluline on vaadata patogeensuse hindamisel mitmeid allikaid.

Tabel 8. Patogeensed variandid lüsosomaalsete ladestushaigustega seotud geenides

Geen	Variant	dbSNP ID või chr:pos	Clin-Var	Inter-Var	Isikute arv (n=4776)	Patogeensuse skoorid
ARSB	c.1450A>G p.Arg484Gly	rs201101343	TP	TP	1	6/8
	c.629A>G p.Tyr210Cys	rs118203943	P	TP	2	8/8
	c.966G>A p.Trp322*	5:78181583	.	P	1	2/2
GAA	c.307T>G p.Cys103Gly	rs398123174	P	TP	1	8/8
	c.1634C>T p.Pro545Leu	rs121907942	P	VUS	1	8/8
	c.1655T>C p.Leu552Pro	rs779556619	P	VUS	2	8/8
	c.1822C>T p.Arg608*	rs749529161	.	P	3	2/2
	c.2238G>C p.Trp746Cys	rs1800312	P	VUS	2	8/8
	c.1552-3C>G	rs375470378	P/TP	.	12	0/2
	c.-32-13T>G	rs386834236	P	.	18	0/2
GALNS	c.1156C>T p.Arg386Cys	rs118204437	P/TP	TP	3	7/8
	c.776G>A p.Arg259Gln	rs118204442	P	TP	1	4/8
GBA	c.1342G>C p.Asp448His	rs1064651	P/TP	P	3	3/8
	c.1226A>G p.Asn409Ser	rs76763715	P/TP	P	11	3/8
	c.754T>A p.Phe252Ile	rs381737	P	P	1	6/8
GLA	c.427G>A p.Ala143Thr	rs104894845	C	TP	2	7/8
GLB1	c.1769G>A p.Arg590His	rs398123351	P	P	1	8/8
	c.1768C>T p.Arg590Cys	rs794727165	P	TP	1	8/8
	c.75+2dupT	rs587776525	P	.	1	0/2
GNS	c.667G>T p.Glu223*	rs3203621	.	P	1	2/2
GUSB	c.1648C>T p.His550Tyr	7:65432723	.	TP	1	8/8
	c.1187G>A p.Cys396Tyr	7:65439570	.	TP	1	8/8
	c.1120C>G p.Arg374Gly	rs747572640	.	TP	2	6/8
	c.959A>C p.Tyr320Ser	rs886044680	TP	TP	2	8/8
	c.658C>G p.Leu220Val	rs760324067	.	TP	3	5/8
	c.530C>T p.Thr177Ile	rs587779400	P	TP	3	4/8

	c.407G>C p.Gly136Ala	rs749073290	.	TP	1	8/8
	c.352A>G p.Arg118Gly	rs767239546	.	TP	1	7/8
	c.307C>T p.Arg103Trp	rs786205673	TP	TP	1	7/8
<i>HGSNAT</i>	c.370A>T p.Arg124Trp	rs754875934	P	VUS	1	6/8
	c.848C>T p.Pro283Leu	rs121908282	P/TP	TP	3	8/8
<i>IDUA</i>	c.178C>T p.Gln60*	4:981616	.	P	1	2/2
	c.208C>T p.Gln70*	rs121965020	P	P	8	2/2
	c.1205G>A p.Trp402*	rs121965019	P	P	3	2/2
	c.1614delG p.His539Thrfs*21	rs727503967	P	.	1	1/2
<i>NAGLU</i>	c.419A>G p.Tyr140Cys	rs753520553	P	TP	1	8/8
	c.889C>T p.Arg297*	rs104894592	P/TP	P	2	2/2
	c.1694G>A p.Arg565Gln	rs104894598	P	TP	1	8/8
<i>NPC1</i>	c.3557G>A p.Arg1186His	rs200444084	P/TP	VUS	1	8/8
	c.3467A>G p.Asn1156Ser	rs28942105	P	VUS	1	8/8
	c.3019C>G p.Pro1007Ala	rs80358257	P	VUS	1	7/8
	c.2974G>C p.Gly992Arg	rs80358254	P/TP	VUS	1	4/8
<i>NPC2</i>	c.190+5G>A	rs80358268	P	.	1	0/2
<i>SGSH</i>	c.1135delG p.Val379Cysfs*34	rs777956287	P	.	1	1/2
	c.892T>C p.Ser298Pro	rs138504221	P	TP	2	7/8
	c.220C>T p.Arg74Cys	rs104894636	P	TP	16	8/8
	c.197C>G p.Ser66Trp	rs104894637	P	TP	1	8/8
<i>SMPD1</i>	c.1826_1828delGCC p.Arg609del	rs769777506	P	.	1	0/2

P – patogeenne

TP – tõenäoliselt patogeenne

C – (*conflicting interpretation*) patogeensuse hinnangutes esineb konflikte (P, TP, VUS)

VUS – ebaselge tähendusega variant (*variant of uncertain significance*)

“.” – info puudub

Tabel 9. LSD liitheterosügoot ja topeltkandjad

Isik	Geen	Variant
Isik 7, liitheterosügoot	<i>GAA</i>	c.2238G>C (p.Trp746Cys)
	<i>GAA</i>	c.1552-3C>G
Isik 8, topeltkandja	<i>GAA</i>	c.1634C>T (p.Pro545Leu)
	<i>GUSB</i>	c.658C>G (p.Leu220Val)
Isik 9, topeltkandja	<i>GBA</i>	c.1226A>G (p.Asn409Ser)
	<i>SGSH</i>	c.220C>T (p.Arg74Cys)
Isik 10, topeltkandja	<i>IDUA</i>	c.208C>T (p.Gln70*)
	<i>GLA</i>	c.427G>A (p.Ala143Thr)

Antud töö käigus leiti 1 liitheterosügootne kandja Pompe tõve suhtes (Tabel 9). Kuna lüsoosomaalsed ladestushaigused on retsessiivse pärandumismustriga, võivad need liitheterosügoots olekus avalduda. Ühes uuringus on leitud, et isikul 7 esineval *GAA* c.2238G>C (p.Trp746Cys) variandil on kergelt patogeenne efekt (Niño jt., 2012), kuid tegemist on kõige levinuma hilise algusega Pompe tõve mutatsioonidest hiinlaste seas (Liu jt., 2014) ning esineb ka HGMD andmebaasis. *GAA* c.1552-3C>G variant mõjutab splaissingut (käsitletud ülal) ning esineb samuti HGMD andmebaasis. Seega võib antud liitheterosügootil olla kõrgeenenud haigusrisk, mille kohta võiks anda tagasisidet. Tegemist on 1970. aastal sündinud naisega, kellel käesoleval hetkel Pompe diagnoosi ei ole. On teada, et Pompe võib avalduda alles täiskasvanueas, kuid hilisvorm on tavaliselt leebem kui imikueas algav vorm.

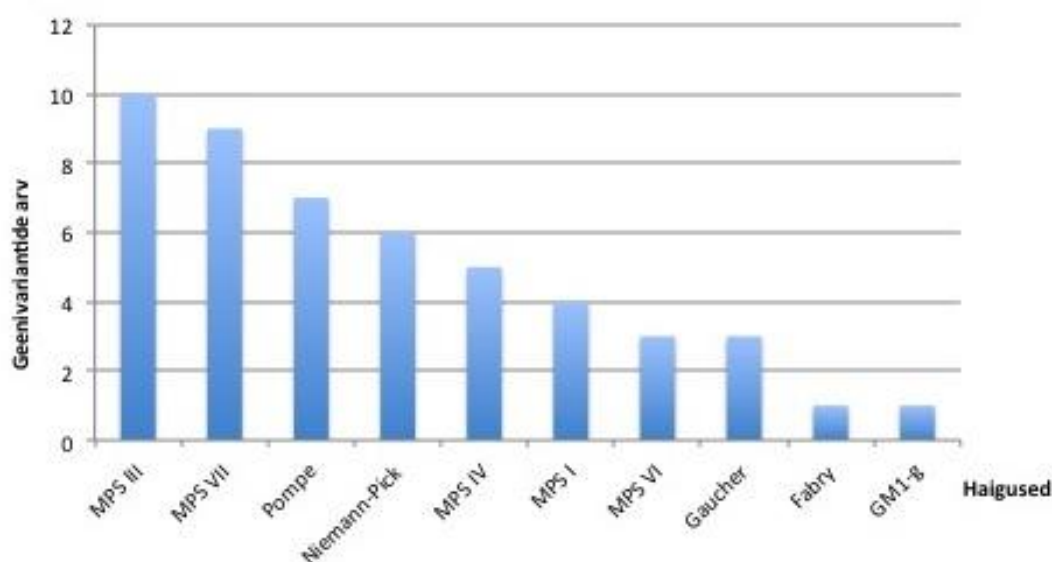
Fabry tõve seoselise geenivariandi kandjaid leiti käesolevas töös 2. On teada, et tegemist on 1952. ja 1966. aastal sündinud naistega. Fabry tõve klassikaline pilt avaldub tavaliselt meestel hemisügoots olekus ning heterosügootse kandlusega naistel reeglina haigus ei avaldu, kuigi ka neil võivad hilisemas eas kerged nähud tekkida. Samuti võib naistel X-kromosoomi inaktivatsiooni tõttu siiski klassikaline Fabry tõbi avalduda, kui inaktiveeritakse normaalne alleel. Seetõttu peaks ka heterosügootseid naisi jälgima (Õunap jt., 2002; Echevarria jt., 2016).

Käesolevas töös olev Fabry tõvega seotud *GLA* geenivariant c.427G>A (p.Ala143Thr) on InterVar andmebaasis hinnatud tõenäoliselt patogeenseks ning ClinVar andmebaasis esineb hinnangutes konflikte (nii patogeenne, tõenäoliselt patogeenne kui ebakindla tähendusega). 8-st kasutatud *in silico* skoorist 7 hindavad geenivarianti patogeenseks. Ühes uuringus (Nance jt., 2006) leiti 34-aastaselt mehel *GLA* geenis Ala143Thr asendus, mis põhjustas plasma ja leukotsüütide α -galaktosidaas A aktiivsuse vähenemist. Tal diagnoositi progresseeruvate jalakrampide ja sellest põhjustatud tugevate valude tõttu hilise algusega Fabry tõbi. Ka tema ema, kes oli antud mutatsiooni osas heterosügootne kandja, koges alates 58. eluaastast samu

sümptomeid. Seega võib antud geenivariant ka käesolevas töös leitud isikuid potentsiaalselt mõjutada.

Üks antud isikutest (isik 10) kannab lisaks ka MPS I seoselist geenivarianti *IDUA* c.208C>T (p.Gln70*). Variant on patogeenne nii ClinVar kui InterVar andmebaasis ning patogeensusust ennustavad 2 skoori 2-st. Samuti esineb variant HGMD ning OMIM andmebaasides. OMIM andmebaasi kirjelduse kohaselt on tegemist variandiga, mis homosügootses olekus põhjustab äärmiselt raskekujulise kliinilise fenotüübi, liitheterosügootses olekus on fenotüübil lai spekter. Antud variandi edasi kandmine võib seega kätkeada endas riski järgmistele põlvkondadele.

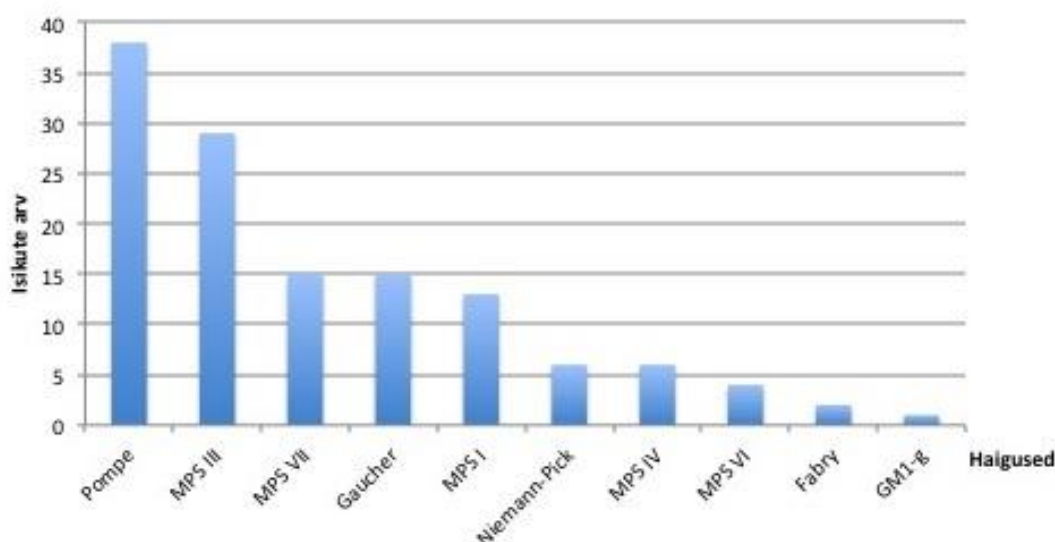
Tulemustest selgus (Joonis 3), et kõige rohkem geenivariante (n=10) esines seoses MPS III-ga geenides *SGSH*, *NAGLU*, *HGSNAT* ja *GNS*. 9 geenivarianti leiti seoses MPS VII-ga geenis *GUSB*. 7 geenivarianti leiti seoses Pompe tõvega geenis *GAA*. 6 geenivarianti leiti seoses Niemann-Pick'i tõvega geenides *SMPD1*, *NPC1*, *NPC2*. 5 geenivarianti leiti seoses MPS IV-ga geenides *GALNS* ja *GLB1* ning 4 geenivarianti seoses MPS I-ga geenis *IDUA*. 3 geenivarianti leiti seoses MPS VI-ga geenis *ARSB*, samuti 3 varianti seoses Gaucher' tõvega geenis *GBA*. 1 geenivariant leiti seoses Fabry tõvega geenis *GLA* ning samuti 1 variant seoses GM1-gangliosidoosiga geenis *GLB1*.



Joonis 3. Lüsosomaalsetele ladestushaiguste seoseliste geenivariantide arv

MPS III sisaldab tüüpe A-D, Niemann-Pick sisaldab tüüpe A/B, C1 ja C2 ning MPS IV sisaldab tüüpe A ja B.

Tulemustest selgus (Joonis 4), et enim indiviide (n=38) kannab Pompe tõve seoselist geenivarianti. Nendest üks isik on liitheterosügoot kahe erineva *GAA* geeni mutatsiooni osas. Seega esineb Pompe tõve seoselist geenivarianti 37 kandjal ja ühel potentsiaalselt kõrge haigusriskiga isikul. Teisel kohal kandjate arvu (n=29) poolest on MPS III, nendest 20 kannavad A tüüpi MPS III seoselist varianti. MPS VII ja Gaucher' tõve seoselisi variante kannab 15 indiviidi, MPS I seoselist varianti 13 indiviidi. MPS IV seoselist varianti kannab 6 isikut ning MPS VI seoselist varianti samuti 6 isikut. Fabry tõve seoselist geenivarianti kannab 2 indiviidi ning GM1-gangliosidoosi seoselist varianti 1 indiviid.



Joonis 4. Lüsosomaalsete ladestushaiguste seoselisi variante kandvate isikute arv

MPS III sisaldab tüüpe A-D, Niemann-Pick sisaldab tüüpe A/B, C1 ja C2 ning MPS IV sisaldab tüüpe A ja B.

Käesolevas töös arvutati tulemuste põhjal lüsosomaalsete ladestushaiguste eeldatav sagedus Eestis ning seda võrreldi kirjanduses avaldatud sagedustega mujal maailmas (Tabel 10). Selgus, et Pompe tõve ja MPS I eeldatavad sagedused vastavad sagedusele kirjanduses. Seevastu Niemann-Pick'i tõve, MPS IV ja MPS VI eeldatavad sagedused erinesid oluliselt kirjanduses seni avaldatud sagedustest. Kuna lüsosomaalsed ladestushaigused on harva esinevad haigused, ei pruugi väikese valimi puhul piisavalt haigusseoselisi indiviide leida. Seetõttu võib potentsiaalselt haigusseoseliste isikute arv olla antud juhtudel tegelikult suurem

kui töös leitud. Seda näitab ka uuring mukopolüsahharidoosidest Eestis perioodil 1985-2006 (Krabbi jt., 2012), kus leiti, et MPS VI sagedus elussündide kohta Eestis on 1/370 000 (käesolevas töös sagedus kandluste põhjal 1/6 250 000). Samuti leiti selles uuringus, et MPS II sagedus on 1/46 287, kuid käesolevas töös ei leitud ühtegi MPS II seoselist geenivarianti kandvat isikut.

Fabry tõve puhul arvutati eeldatav sagedus naiste puhul, kuna ühtegi Fabry tõve seoselist geenivarianti meestel ei leitud. Tavaliselt avalduvad X-liitelised haigused meestel ning seetõttu loetakse üldiselt sellistel juhtudel haiguse sagedust meeste kohta. Näiteks Eestis on Fabry esinemise sageduseks ennustatud 10 - 15 meest elanikkonna kohta (Õunap jt., 2002).

Tabel 10. Lüsosomaalsete ladestushaiguste eeldatav sagedus Eestis võrdluses kirjanduses avaldatud sagedustega mujal maailmas

Haigus	Sagedus	Sagedus kirjanduses ²
Pompe tõbi	1/56 600	1/60 000 - 100 000
Fabry tõbi	1/1130*	1/3000 - 100 000
Gaucher' tõbi	1/390 600	1/60 000 - 1/100 000
Niemann-Pick'i tõbi	1/2 770 000	1/250 000 - 1/500 000
MPS I	1/510 200	1/100 000 - 1/500 000
MPS III	1/104 000	1 /70 000
MPS IV	1/1 562 500	1/250 000 - 1/926 000
MPS VI	1/6 250 000	1/250 000 - 1/600 000
MPS VII	1/390 600	1/250 000

² Sageduse andmed põhinevad GeneReviews® (<https://www.ncbi.nlm.nih.gov/books/NBK1116/>); Orphanet (<https://www.orpha.net/>); <https://ghr.nlm.nih.gov/condition>.

*Fabry tõve puhul on tegemist X-liitelise haigusega ning antud tulemuste põhjal on välja toodud sagedus naiste seas.

Käesoleva töö käigus leiti lüsosomaalsete ladestushaigustega seoses 48 patogeenset või tõenäoliselt patogeenset geenivarianti 15 geenis. Leitud geenivariandid esinevad kokku 126 indiviidil 4776-st. Tulemustest selgus, et lüsosomaalsete ladestushaigustega seotud andmebaaside patogeensuse hinnangud läksid üldiselt *in silico* skooride hinnangutega kokku ning erinevusi esines vaid mõnel juhul. Siiski tuleb patogeensuse hinnangutel lähtuda mitmetest allikatest. Tulemustest selgus ka, et kõige rohkem geenivariante leiti seoses MPS III ja VII tüüpi haigustega. Seevastu kõige enam võiks Eestis esineda Pompe tõbe, mille sagedus vastab ka maailmas esinevale sagedusele. Sageduselt teine LSD Eestis on tulemuste järgi MPS IIIA. Varasemalt läbi viidud mukopolüsahharidooside uuringu põhjal on MPS IIIA

teine kõige levinum mukopolüsahharidoos Eestis (Krabbi jt., 2012). Käesolevas töös leiti, et kõige vähem levinud on MPS VI, mis kuulub vähem levinud LSD-de hulka enamikus populatsioonides (Krabbi jt., 2012). Kuna tegemist on autosoom-retsessiivsete haigustega, ei ole plaanis kandluste kohta tulevikus isikutele tagasisidet anda. Kuid töös tuvastati üks Pompe tõve liitheterosügoot, kellel on eeldatavalt kõrge haigusrisk ning kes võib olla tulevikus potentsiaalne tagasiside saaja. Fabry haiguse heterosügootne kandlus naiste puhul võiks olla kasulik arstidele juhul, kui hilise algusega mutatsioon neil naistel avalduda võib. Kuna tegemist on haruldase haigusega, oleks arstidel mutatsiooni esinemise põhjal kergem haigust diagnoosida.

Oluline on mõista, et patogeensete geenivariantide esinemine ja sellest tulenev kõrge haigusrisk võib küll suurendada haiguse avaldumise tõenäosust, kuid ei tähenda tingimata haiguse avaldumist. Võib esineda geenivariandi mittetäielik penetrantsus, mille puhul kõigil mutatsioonidega isikutel kliinilisi tunnuseid ei teki. Selles võivad rolli mängida mitmed tegurid, sh eluviis ja keskkond, aga ka teised geneetilised faktorid. Haiguse avaldumist võivad mõjutada ka vanus ja sugu.

Enne potentsiaalset tagasiside andmist tuleks eelnevalt uuesti üle kontrollida variandi patogeensus, kuna patogeensususe hinnangud võivad aja jooksul vähem uuritud geenivariantide puhul muutuda ning seejärel valideerida leidude esinemine (nt Sangeri meetodil).

Antud töö käigus uuriti 4776 geenidoonoril sekkumistõhusate geenivariantide esinemist ning leiti 120 indiviidi, kellele potentsiaalselt tagasisidet anda. Sekkumistõhusate nimekirjade alusel leiti, milliseid haigusi antud nimekirjadest võib Eestis kõige rohkem esineda. Tänu suurtele WGS ja WES andmemassiividele on võimalik tuvastada inimestel harvaesinevaid haigusi, mis võiksid jääda muidu diagnoosimata ning aidata ennetades või õigeaegse raviga suurendada inimeste tervena elatud aastate arvu. Seetõttu on populatsioonipõhistest biopangadest saadav informatsioon olulise väärtusega ning seda peaks tulevikus laialdasemalt ära kasutama.

KOKKUVÕTE

Käesoleva magistritöö käigus analüüsiti TÜ Eesti Geenivaramu 4776 geenidoonori eksoomi ja täisgenoomi sekveneerimisel saadud andmetest haigusseoselisi geenivariante ACMG ja Geisinger-76 sekkumist võimaldavate nimekirjade 78 geenis, ning lisaks 16 lüsosomaalseid ladestushaigusi põhjustavas geenis. Geenivariantide haigusseoselisuse hindamiseks vaadeldi alleelisagedusi, *in silico* ennustuslikke skooride ning kliiniliste andmebaaside hinnanguid. Töö käigus arvutati ka geenivariantidele vastavate haiguste eeldatav sagedus. Töö tulemusena leiti ACMG/GHS76 nimekirja alusel 77 patogeenset või tõenäoliselt patogeenset geenivarianti 36 geenis ja lüsosomaalsete ladestushaigustega seoses 48 varianti 15 geenis. Üldiselt läksid skooride patogeensuse ennustused andmebaaside hinnangutega kokku. Geenivariandid esinesid vastavalt 237 ja 126 indiviidil. Tulemustest selgus, et ACMG/GHS76 nimekirja alusel võiks potentsiaalselt kõrge haigusriskiga seotud olla 119 indiviidi, lüsosomaalsete ladestushaiguste riskiga 1 *GLA* geeni liitheterosügootse kandlusega indiviid. Tulemuste alusel võiks nimekirjades olevatest haigustest Eestis kõige enam esineda pärilikku rinna- ja munasarjavähki, maliigse hüpertermia eelsoodumust, hüpertroofilist ja dilatatiivset kardiomüopaatiat ning pika QT sündroomi. Lüsosomaalsetest ladestushaigusest võiks kõige enam esineda Pompe tõbe ning III tüüpi mukopolüsahharidoosi.

Kokku võib 120 inimesel esineda potentsiaalselt kõrge haigusrisk, mille puhul oleks võimalik ennetuse või varajase raviga haigusest tulenevaid komplikatsioone ära hoida või vähendada. Seetõttu tuleks järgnevalt antud leiud valideerida ning kaaluda personaliseeritud genoomikapõhist tagasisidet.

Populatsioonipõhised biopangad võimaldavad kaasajal suurte geneetiliste andmehulkade tõttu tuvastada indiviidide suurenenud haigusriski ja kui rakendada need teadmised meditsiini võib pikendada inimeste tervena elatud aastaid.

Actionable gene variants in Estonian Genome Center sample – identification of disease-associated variants and their possible use in genetic feedback

Karmen Vaiküll

Summary

As whole genome and exome sequencing are being more widely used in human genetics, it is possible to find more high risk genetic variants that are clinically relevant. Thus there has been increasing attention to the return of genomic results to get the diagnosis and a suitable treatment at the right time. For that purpose there are created actionable gene lists and many computational tools and clinical databases that help to assess the gene variant pathogenicity.

During this study whole-genome and whole-exome sequencing data was analyzed in 4776 gene donors of Estonian Genome Center of University of Tartu to search for actionable gene variants in 78 genes that are named in American College of Medical Genetics and Genomics (ACMG) and Geisinger Health System (Geisinger) actionable gene lists. Also 16 genes associated with lysosomal storage diseases were evaluated. As the treatment is developing, these diseases become actionable.

In this study there were found 77 disease-associated gene variants in 36 genes of ACMG/Geisinger lists and 48 variants in 15 genes associated with lysosomal storage diseases. These gene variants appeared in 237 and 126 individuals respectively. A lot of known disease-associated variants were associated with autosomal-recessive diseases without any alternative homozygotes found. Individuals carrying those variants were not considered to receive possible feedback. Including one *GLA* gene compound heterozygote, 120 individuals could possibly have a high disease risk and should receive genetic feedback.

Also based on the data found in the study of the diseases assessed, hereditary breast and ovarian cancer, long QT syndrome, malignant hyperthermia susceptibility, hypertrophic and dilatative cardiomyopathy could be most frequent in Estonia from ACMG/Geisinger lists. From lysosomal storage diseases Pompe disease and mucopolysaccharidosis type III could be most frequent.

Returning genomic results to participants could help to avoid complications due to diseases, so data obtained in population-based biobanks have a great value and should be used more widely to increase the number of healthy life years of individuals.

TÄNUAVALDUSED

Soovin tänada oma juhendajaid Tiit Nikopensiust ja Neeme Tõnissoni võimaluse eest nendega koostööd teha.

Minu eriline tänu kuulub ka Marili Paloverile, kes mind erinevates töö etappides aitas.

Lisaks soovin tänada Mart Kalsi annotatsioonide eest ning kõiki TÜ Eesti Geenivaramu töötajaid, kes töö valmimisele oma panuse andsid.

KASUTATUD KIRJANDUSE LOETELU

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A. S., Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods*. 7(4):248–249.

Alders, M., Bikker, H., Christiaans, I. (2003, [Last updated 2018]). Long QT Syndrome. In: Pagon, R. A., Adam, M. P., Ardinger, H. H., et al., editors. *GeneReviews®* [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK1129/>

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A. F., Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 43(D1):D789–798.

Aretz, S., Tricarico, R., Papi, ... Genuardi, M. (2014). MUTYH-associated polyposis (MAP): Evidence for the origin of the common European mutations p.Tyr179Cys and p.Gly396Asp by founder events. *Eur. J. Hum. Genet*. 22(7):923–929.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., ... Wilson, R. K. (2015). A global reference for human genetic variation. *Nature*. 526(7571):68–74.

Beck, M. (2018). Treatment strategies for lysosomal storage disorders. *Dev. Med. Child. Neurol*. 60(1):13–18.

Bijlsma, R.M., Wouters, R.H.P., Wessels, H., May, A.M., Ausems, M.G.E.M., Voest, E.E., Bredenoord, A.L. (2018). Managing unsolicited findings in genomics: a qualitative interview study with cancer patients. *Psychooncology*. 27(4):1327-1333.

Blackburn, H.L., Schroeder, B., Turner, C., Shriver, C.D., Ellsworth, D.L., Ellsworth, R.E. (2015). Management of Incidental Findings in the Era of Next-generation Sequencing. *Curr. Genomics*. 16(3):159–174.

Chakraborty, S., Rafi, M. A., Wenger, D. A. (1994). Mutations in the Lysosomal , □-Galactosidase Gene That Cause the Adult Form of GM I Gangliosidosis. *Am. J. Hum. Genet*. 54:1004-1013.

Crehalet, H., Millat, G., Albuissou, J., Bonnet, V., Rouvet, I., Rousson, R., Bozon, D. (2012). Combined use of in silico and in vitro splicing assays for interpretation of genomic variants of unknown significance in cardiomyopathies and channelopathies. *Cardiogenetics*. 2(1):6.

- den Dunnen, J. T., Dalglish, R., Maglott, D. R., Hart, R. K., Greenblatt, M.S., McGowan-Jordan, J., Roux A. F., Smith, T., Antonarakis, S. E., Taschner, P. E. M. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* 37(6):564–569.
- Dewey, F. E., Murray, M. F., Overton, J.D., ... Carey, D. J. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science.* 354(6319):aaf6814-1- aaf6814-10.
- Dorschner, M. O., Amendola, L. M., Turner, E. H., Jarvik, G. P. (2013). Actionable, Pathogenic Incidental Findings in 1,000 Participants' Exomes. *Am. J. Hum. Genet.* 93(4):631–640.
- Durbin, R. M., Altshuler, D., Abecasis, G. R., ... Francis S. Collins. (2010). A map of human genome variation from population scale sequencing. *Nature.* 476(7319):1061–1073.
- Echevarria, L., Benistan, K., Toussaint, A., Dubourg, O., Hagege, A. A., Eladari, D., Jabbour, F., Beldjord, C., De Mazancourt, P., Germain, D. P. (2016). X-chromosome inactivation in female patients with Fabry disease. *Clin. Genet.* 89(1):44–54.
- Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Hum. Mutat.* 32(5):557–563.
- Fossey, R., Kochan, D., Winkler, E., ... Kullo, I. J. (2018). Ethical considerations related to return of results from genomic medicine projects: The eMERGE network (phase III) experience. *J. Pers. Med.* 8(1):1–18.
- González-Pérez, A., López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet.* 88(4):440–449.
- Green, R. C., Berg, J.S., Grody, W.W., ... Biesecker, L. G. (2013). ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. *Genet. Med.* 15(7):565–574.
- Hu, D., Barajas-Martinez, H., Burashnikov, E., Antzelevitch, C. (2009). A mutation in the $\beta 3$ subunit of the cardiac sodium channel associated with brugada ECG phenotype. *Circ. Cardiovasc. Genet.* 2(3):270–278.

- Johns, A.L., McKay, S.H., Humphris, J.L., ... Biankin, A.V. (2017). Lost in translation: Returning germline genetic results in genome-scale cancer research. *Genome Med.* 9(1):1–10.
- Johnson, K.J., Gehlert, S. (2014). Return of results from genomic sequencing: A policy discussion of secondary findings for cancer predisposition. *J Cancer Policy.* 2(3):75–80.
- Kalia, S. S., Adelman, K., Bale, S. J., ... Miller, D.T. (2016). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 19(2):249–255.
- Kantorovich, V., King, K. S., Pacak, K. (2011). SDH-related Pheochromocytoma and paraganglioma. *Best Pract. Res. Clin. Endocrinol. Metab.* 24(3):415–424.
- Kircher, M., Witten, D. M., Jain, P., O’roak, B. J., Cooper, G. M., Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 46(3):310–315.
- Kostareva, A., Kiselev, A., Gudkova, A., ... Shlyakhto, E. (2016). Genetic spectrum of idiopathic restrictive cardiomyopathy uncovered by next-generation sequencing. *PLoS One.* 11(9):1–16.
- Krabbi, K., Joost, K., Zordania, R., Talvik, I., Rein, R., Huijmans, J. G. M., Verheijen, F. V., Õunap, K. (2012). The Live-Birth Prevalence of Mucopolysaccharidoses in Estonia. *Genet. Test. Mol. Biomarkers.* 16(8):846–849.
- Kroos, M., Manta, P., Mavridou, I., Muntoni, F., Halley, D., Van der Helm R., Zaifeiriou, D., Van der Ploeg, A., Reuser, A., Michelakakis, H. (2006). Seven cases of Pompe disease from Greece. *J. Inherit. Metab. Dis.* 29(4):556–563.
- Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications, Next Generation Sequencing Jerzy Kulski, IntechOpen, DOI: 10.5772/61964
- Kumar, P., Henikoff, S., Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4(7):1073–1082.
- Landrum, M. J., Lee, J. M., Benson, M., Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44(D1):D862-868.
- Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25:1754–1760.

- Li, Q., Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* 100(2):267–280.
- Liu, X., Wang, Z., Jin, W., Lv, H., Zhang, W., Que, C., Huang, Y., Yuan, Y. (2014). Clinical and GAA gene mutation analysis in mainland Chinese patients with late-onset Pompe disease: identifying c.2238G > C as the most common mutation. *BMC Med. Gen.* 15:141-149.
- Liu, X., Wu, C., Li, C., Boerwinkle, E. (2015). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* 37:235–241.
- Mavridou, I., Dimitriou, E., Vanier, M. T., Vilageliu, L., Grinberg, D., Latour, P., Xaidara, A., Lycopoulou, L., Bostantjopoulou, S., Zafeiriou, D., Michelakakis, H. (2017). The Spectrum of Niemann-Pick Type C Disease in Greece. *JIMD Rep.* 36:41-48.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17:122-135.
- Murtagh, M.J., Demir, I., Harris, J.R., Burton, P.R. (2011). Realizing the promise of population biobanks: A new model for translation. *Hum Genet.* 130(3):333–345.
- Nance, C. S., Klein, C. J., Banikazemi, M., ... Desnick, R. J. (2006). Later-Onset Fabry Disease. *Arch. Neurol.* 63:453-457.
- Niño, M. Y., Mateus, H. E., Fonseca, D. J., Kroos, M. A., Ospina, S. Y., Mejía, J. F., Uribe, J. A., Reuser, A. J. J., Laissue, P. (2012). Identification and Functional Characterization of GAA Mutations in Colombian Patients Affected by Pompe Disease. *JIMD Rep.* 7: 39–48.
- Pajusalu, S., Talvik, I., Noormets, K., ... Reimand, T. (2016). De novo exonic mutation in MYH7 gene leading to exon skipping in a patient with early onset muscular weakness and fiber-type disproportion. *Neuromuscul. Disord.* 26(3):236–239.
- Parodi, B. (2015) Biobanks: A Definition. In: Mascalzoni D. (eds) *Ethics, Law and Governance of Biobanking. The International Library of Ethics, Law and Technology*, vol 14. Springer, Dordrecht. 14:15–20.
- Petrucelli, N., Daly, M. B., Pal, T. (1998 [Last updated 2016]). BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. In: Pagon, R. A., Adam, M. P., Ardinger, H. H., et al., editors. *GeneReviews®* [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK1247/>

- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20(1):110–121.
- Pulford, D.J., Harter, P., Floquet, A., ... du Bois, A. (2016). Communicating BRCA research results to patients enrolled in international clinical trials: Lessons learnt from the AGO-OVAR 16 study. *BMC Med. Ethics.* 17(1):1–11.
- Reva, B., Antipin, Y., Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* 39(17):37–43.
- Richards, S., Aziz, N., Bale, S., ... Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 17(5):405–423.
- Rigante, D., Cipolla, C., Basile, U., Gulli, F., Savastano, M.C. (2017). Overview of immune abnormalities in lysosomal storage disorders. *Immunol Lett.* 188:79–85.
- Sanderson, S.C., Linderman, M.D., Suckiel, S.A., Zinberg, R., Wasserstein, M., Kasarskis, A., George, A. D, Schadt, E. E. (2017). Psychological and behavioural impact of returning personal results from whole-genome sequencing: The HealthSeq project. *Eur. J. Hum. Genet.* 25(3):280–292.
- Schultz, M.L., Tecedor, L., Chang, M., Davidson, B.L. (2011). Clarifying lysosomal storage diseases. *Trends Neurosci.* 34(8):401–410.
- Schwarz, J. M, Rödelberger, C., Schuelke, M., Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods.* 7(8):575–576.
- Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29(1):308–311.
- Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A. D., Cooper, D. N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* 136(6):665–677.
- Swede, H., Stone, C.L., Norwood, A.R. (2007) National population-based biobanks for genetic research. *Genet Med.* 9(3):141–149.

- Tamboom, K., Kaasik, K., Aršavskaja, J., Tekkel, M., Lilleorg, A., Padrik, P., Metspalu, A., Veidebaum, T. (2010). BRCA1 mutations in women with familial or early-onset breast cancer and BRCA2 mutations in familial cancer in Estonia. *Hered. Cancer Clin. Pract.* 8(1):4–10.
- Waldmüller, S., Müller, M., Rackebrandt, K., Binner, P., Poths, S., Bonin, M., Scheffold, T. (2008). Array-based resequencing assay for mutations causing hypertrophic cardiomyopathy. *Clin. Chem.* 54(4):682–687.
- Wang, R. Y., Bodamer, O. A., Watson, M.S., Wilcox, W.R. (2011). Lysosomal storage diseases: Diagnostic confirmation and management of presymptomatic individuals. *Genet Med.* 13(5):457–484.
- Õunap, K., Anijalg, E., Pikk, O., Kukk, T. (2002). Pärilik lüsoosomaalne ladestumishaigus – Fabry tõbi. *Eesti Arst.* 81(2):77–84.
- Xu, H., Ren, D. (2016). Lysosomal Physiology. *Annu. Rev. Physiol.* 77: 57–80.
- Yang, H., Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 10:1556–1566.

KASUTATUD VEEBIAADRESSID

<https://www.geisinger.org/mycode>

<https://mpssociety.org/learn/diseases/mps-vii/>

<http://www.hgmd.org>

<https://www.ncbi.nlm.nih.gov/clinvar/>

<http://wintervar.wglab.org/>

<http://www.omim.org>

<http://www.ncbi.nlm.nih.gov/SNP/>

<http://www.lovd.nl>

<http://gnomad.broadinstitute.org/>

<http://evs.gs.washington.edu/EVS/>

<http://cadd.gs.washington.edu;>

<http://sift.bii.a-star.edu.sg;>

<http://mutationassessor.org;>

<http://genetics.bwh.harvard.edu/pph2>

<http://www.mutationtaster.org>

<http://bbglab.irbbarcelona.org/fannsdb>

<http://annovar.openbioinformatics.org/en/latest/user-guide/filter/>

<https://wpcalc.com/en/equilibrium-hardy-weinberg/>

<https://www.kliinikum.ee/geneetikakeskus/molekulaardiagnostika/48-wilsoni-tobi>

<https://www.ncbi.nlm.nih.gov/books/NBK11116/>

<https://www.orpha.net/>

<https://www.uptodate.com/contents/susceptibility-to-malignant-hyperthermia-evaluation-and-management>

<https://ghr.nlm.nih.gov/condition>

LISA1 ACMG/Geisinger nimekirja geenivariantide *in silico* skoorid

Geen	Variant	SIFT	pp2_hdiv	pp2_hvar	Condel	mt	CADD_Phred	PhyloP	ma
<i>ACVRL1</i>	c.920C>T p.Ala307Val	0.14	0.995	0.676	N	D	23.8	2.924	M
<i>APC</i>	c.481C>T p.Gln161*	0	.	.	.	A	38	5,638	.
<i>APOB</i>	c.10898G>A p.Trp3633*	1	.	.	.	D	48	3.111	.
<i>ATP7B</i>	c.4135C>T p.Pro1379Ser	0	1.0	0.998	D	D	28.6	2.912	L
	c.3895C>G p.Leu1299Val	0	1.0	0.998	D	D	25.1	3.045	M
	c.3818C>T p.Pro1273Leu	0	1.0	1.0	D	D	29.7	5.832	M
	c.3505A>G p.Met1169Val	0.01	0.99	0.975	D	D	26.2	4.928	N
	c.3207C>A p.His1069Gln	0	1.0	1.0	D	A	25.1	0.543	H
	c.2605G>A p.Gly869Arg	0	1.0	1.0	D	D	34	6.198	H
	c.2532delA p.Val845Serfs *28	33	6.202	.
	c.1630C>T p.Gln544*	0.39	.	.	.	D	36	1.749	.
	c.658G>T p.Gly220*	0.1	.	.	.	A	38	3.884	.
	c.5329_5330insC p.Gln1777Profs *74	35	-0,516	.
<i>BRCA1</i>	c.4258C>T p.Gln1420*	0	.	.	.	A	40	2,421	.
	c.4065_4068delTCAA p.Asn1355Lysfs *10	29,2	0,284	.
	c.4035delA p.Glu1346Lysfs *20	21,5	2,074	.
	c.1840A>T p.Lys614*	1	.	.	.	A	36	1,002	.
	c.1687C>T p.Gln563*	0,02	.	.	.	A	35	1,122	.
	c.9381G>A p.Trp3127*	1	.	.	.	A	49	2,555	.
<i>BRCA2</i>	c.37G>T p.Glu13*	0,01	.	.	.	A	38	3,458	.
	c.6402_6406delTAACT p.Asn2135Lysfs *3	28,5	0,208	.
	c.8572C>T p.Gln2858*	0,09	.	.	.	A	47	2,337	.
	c.2707C>T p.Arg903*	1	.	.	.	A	39	0.841	.
<i>CACNA1S</i>	c.502C>T	1	.	.	.	A	39	2.955	.

	p.Arg168*								
CAV3	c.277G>A p.Ala93Thr	0	1.0	0.999	D	D	33	5.03	M
COL3A1	c.1475G>A p.Gly492Glu	0	1.0	1.0	.	D	26.1	5.268	H
	c.1714C>T p.Arg572*	1	.	.	.	A	40	2.501	.
DSG2	c.2315T>G p.Leu772*	1	.	.	.	D	42	3.739	.
DSP	c.4357C>T p.Gln1453*	0.97	.	.	.	D	40	3.053	.
FBN1	c.7339G>A p.Glu2447Lys	0	0.999	0.993	D	D	35	6.179	H
GLA	c.427G>A p.Ala143Thr	0.02	1.000	0.922	D	D	29.9	4.418	.
KCNE2	c.79C>T p.Arg27Cys	0	1.0	0.999	D	D	29.3	2.067	L
	c.161T>C p.Met54Thr	0	0.959	0.551	D	D	24.8	2.521	L
KCNH2	c.2587C>T p.Arg863*	1	.	.	.	D	43	1.366	.
KCNQ1	c.477+1G>A	D	24.2	3.944	.
	c.604G>A p.Asp202Asn	0	1.0	0.999	D	D	34	5.346	H
	c.1664G>A p.Arg555His	0	1.0	1.0	D	D	34	4.802	M
	c.683+5G>A	11.42	0.556	.
LDLR	c.643C>T p.Arg215Cys	0	0.999	0.978	D	D	28.1	0.619	H
	c.768C>A p.Asp256Glu	0	0.907	0.899	D	D	24.4	3.964	H
	c.1202T>A p.Leu401His	0	1.0	1.0	D	D	25.5	4.546	M
	c.1335C>A p.Asp445Glu	0	0.984	0.976	D	D	28.3	1.405	H
MLH1	c.1668-1G>T	0	.	.	.	D	27	5,546	.
MSH2	c.2131C>T p.Arg711*	1	.	.	.	A	50	3,489	.
MSH6	c.1095G>A p.Trp365*	0,98	.	.	.	D	38	5,593	.
	c.3226C>T p.Arg1076Cys	0,1	0,991	0,818	D	D	32	1,894	M
	c.742C>G p.Arg248Gly	0,01	0,943	0,535	D	D	24,5	1,364	M
MUTYH	c.725G>A p.Arg242His	0	1	0,994	D	D	33	5,624	M
	c.316C>T p.Arg106Trp	0	1	1	D	D	27,4	0,935	H
	c.527A>G p.Tyr176Cys	0	1	1	D	A	24,7	4,557	H
	c.1178G>A p.Gly393Asp	0,01	1	0,999	D	A	29,4	5,618	H
MYH7	c.2348G>A p.Arg783His	0.02	0.771	0.514	N	N	25.2	0.671	L
	c.5655G>A p.Ala1885=	15.13	0.77	.

<i>MYL2</i>	c.401A>C p.Glu134Ala	0	0.785	0.605	D	D	28.6	4.291	H
<i>MYL3</i>	c.461G>A p.Arg154His	0	0.999	0.861	N	A	26.8	5.042	M
	c.262A>G p.Thr88Ala	0	0.991	0.853	D	D	25.5	4.21	H
	c.254A>G p.Gln85Arg	0.01	0.961	0.508	D	D	25.4	2.606	M
	c.170C>A p.Ala57Asp	0	1.0	0.999	D	D	29.8	5.096	H
<i>PMS2</i>	c.861_864del ACAG p.Arg287Serfs *19	35	3,293	.
<i>RET</i>	c.2410G>A p.Val804Met	0	1.0	1.0	D	D	26.7	6.018	N
<i>RYR1</i>	c.325C>T p.Arg109Trp	0	1.0	0.999	D	A	34	2.579	M
	c.1840C>T p.Arg614Cys	0.01	1.0	0.999	D	D	35	2.936	M
	c.7268T>A p.Met2423Lys	0	0.96	0.948	N	A	24.0	4.371	M
	c.9579C>G p.Cys3193Trp	0	1.0	0.997	N	D	24.9	1.9	M
	c.14385G>A p.Trp4795*	1	.	.	.	A	58	5.716	.
<i>RYR2</i>	c.3151C>T p.Arg1051Cys	0.28	0.999	0.905	N	D	23.2	0.144	L
<i>PKP2</i>	c.2146-1G>C	D	25.8	3.866	.
<i>SCN3B</i>	c.29T>C p.Leu10Pro	0.06	0.026	0.04	N	A	21.3	0.6	L
<i>SCN5A</i>	c.5872C>T p.Arg1958*	1	.	.	.	D	36	1.135	.
	c.3691G>A p.Glu1231Lys	0.11	0.146	0.053	N	D	23.4	2.101	N
	c.2911C>T p.Arg971Cys	0.03	0.995	0.685	N	D	22.5	1.077	L
	c.655C>T p.Arg219*	1	.	.	.	A	37	2.944	.
<i>SDHC</i>	c.405+1G>T	25.4	4.159	.
<i>STK11</i>	c.368A>G p.Gln123Arg	0,12	0,958	0,776	D	D	27,7	4,273	N
<i>TNNI3</i>	c.422G>A p.Arg141Gln	0	0.975	0.534	D	D	34	1.748	M
<i>TNNT2</i>	c.780+5G>A	11.64	2.3	.
<i>TP53</i>	c.542G>A p.Arg181His	0	1	0,966	D	D	28,5	1,419	M

pp2_hdiv – Polyphen 2 HDiv; pp2_hvar – Polyphen 2 HVar; mt – MutationTaster;

ma – MutationAssessor; “.” – info puudub.

Rasvases kirjas on toodud variandi kahjulikku efekti ennustavad hinnangud.

LISA 2 Lüsosomaalsete ladestushaiguste seoseliste geenivariantide *in silico* skoorid

Geen	Variant	SIFT	pp2_hdiv	pp2_hvar	Condel	mt	CADD_Phred	PhyloP	ma
ARSB	c.1450A>G p.Arg484Gly	0.003	0.964	0.843	D	D	25.7	1.467	H
	c.629A>G p.Tyr210Cys	0.0	1.0	1.0	D	A	27.5	4.417	H
	c.966G>A p.Trp322*	A	42	5.841	.
GAA	c.307T>G p.Cys103Gly	0.0	1.0	1.0	D	D	23.9	4.569	H
	c.1634C>T p.Pro545Leu	0.0	1.0	1.0	D	A	27,8	5.981	M
	c.1655T>C p.Leu552Pro	0.0	1.0	0.995	D	D	25.9	4.261	H
	c.1822C>T p.Arg608*	A	38	3.815	.
	c.2238G>C p.Trp746Cys	0.001	1.0	1.0	D	D	33	5.369	M
	c.1552-3C>G	14.15	2.504	.
	c.-32-13T>G	4.589	-0.117	.
GALNS	c.1156C>T p.Arg386Cys	0.002	1.0	0.994	D	A	33	4.913	L
	c.776G>A p.Arg259Gln	0.275	0.982	0.642	N	A	24.5	3.595	N
GBA	c.1342G>C p.Asp448His	0.069	0.062	0.098	N	A	22.7	4.678	M
	c.1226A>G p.Asn409Ser	0.013	0.368	0.292	N	D	22.7	3.86	L
	c.754T>A p.Phe252Ile	0.002	0.009	0.077	N	A	23.6	3.61	M
GLA	c.427G>A p.Alala143Thr	0.02	1.000	0.922	D	D	29.9	4.418	.
GLB1	c.1769G>A p.Arg590His	0.0	1.0	1.0	D	D	35	5.916	H
	c.1768C>T p.Arg590Cys	0.0	1.0	1.0	D	D	35	5.916	H
	c.75+2dupT	11.72	0.73	.
GNS	c.667G>T p.Glu223*	A	38	2.046	.
GUSB	c.1648C>T p.His550Tyr	0.0	0.997	0.973	D	D	25.5	5.801	M
	c.1187G>A p.Cys396Tyr	0.002	1.0	0.996	D	D	26	3.75	H
	c.1120C>G p.Arg374Gly	0.003	0.986	0.661	D	D	28,5	2.213	M
	c.959A>C p.Tyr320Ser	0.0	1.0	1.0	D	D	24.8	4.636	M
	c.658C>G p.Leu220Val	0.018	0.611	0.414	N	D	23.6	3.665	M
	c.530C>T p.Thr177Ile	0.088	0.991	0.893	N	D	23.9	2.7	M
	c.407G>C p.Gly136Ala	0.001	1.0	0.999	D	D	24.5	5.244	H
	c.352A>G p.Arg118Gly	0.001	1.0	0.999	D	D	23.5	0.251	H
	c.307C>T	0.0	0.595	0.079	D	D	24.7	0.815	M

	p.Arg103Trp								
<i>HGSNAT</i>	c.370A>T p.Arg124Trp	0.006	0.997	0.914	D	N	24.5	0.136	M
	c.848C>T p.Pro283Leu	0.002	1.0	1.0	D	A	34	3.713	H
<i>IDUA</i>	c.178C>T p.Gln60*	A	36	0.926	.
	c.208C>T p.Gln70*	A	35	1.937	.
	c.1205G>A p.Trp402*	A	38	1.628	.
	c.1614delG p.His539Thrfs *21	29.2	2.163	.
<i>NAGLU</i>	c.419A>G p.Tyr140Cys	0.0	1.0	1.0	D	D	27.4	4.422	H
	c.889C>T p.Arg297*	A	36	0.488	.
	c.1694G>A p.Arg565Gln	0.0	1.0	0.999	D	D	32	5.38	H
<i>NPC1</i>	c.3557G>A p.Arg1186His	0.0	1.0	0.999	D	D	33	6.259	H
	c.3467A>G p.Asn1156Ser	0.0	1.0	0.998	D	A	26.3	5.154	M
	c.3019C>G p.Pro1007Ala	0.0	0.995	0.882	D	A	27,7	6.107	M
	c.2974G>C p.Gly992Arg	0.466	0.883	0.444	N	D	25,7	4.362	M
<i>NPC2</i>	c.190+5G>A	22.1	5.577	.
<i>SGSH</i>	c.1135delG p.Val379Cysfs *34	34	4.524	.
	c.892T>C p.Ser298Pro	0.017	1.0	0.997	D	D	25.5	4.587	L
	c.220C>T p.Arg74Cys	0.0	1.0	1.0	D	A	34	3.494	H
	c.197C>G p.Ser66Trp	0.0	1.0	0.996	D	A	32	3.562	M
<i>SMPD1</i>	c.1826_1828d elGCC p.Arg609del	20.8	3.765	.

pp2_hdiv – Polyphen 2 HDiv; pp2_hvar – Polyphen 2 HVar; mt – MutationTaster;
ma – MutationAssessor; “.” – info puudub.

Rasvases kirjas on toodud variandi kahjulikku efekti ennustavad hinnangud.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Karmen Vaiküll
(sünnikuupäev: 20.05.1994)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

“Sekkumist võimaldavad geenivariandid TÜ Eesti Geenivaramu valimis – haigusseoselisuse tuvastamine ja võimalik kasutamine geneetilises tagasisides”,

mille juhendajad on Tiit Nikopensius ja Neeme Tõnisson,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 28.05.2018